Statistical approaches to explore clinical heterogeneity in psychosis

Islam, Atiqul

Statistical approaches to explore clinical heterogeneity in psychosis

Md. Atiqul Islam

Uitgave in de RGOc-reeks, nummer 37.

The infrastructure for the GROUP study is funded through the Geestkracht programme of the Dutch Health Research Council (Zon-Mw, grant number 10-000-1001), and matching funds from participating pharmaceutical companies (Lundbeck, AstraZeneca, Eli Lilly, Janssen Cilag) and universities and mental health care organizations (Amsterdam: Academic Psychiatric Centre of the Academic Medical Center and the mental health institutions: GGZ Ingeest, Arkin, Dijk en Duin, GGZ Rivierduinen, Erasmus Medical Centre, GGZ Noord Holland Noord. Groningen: University Medical Center Groningen and the mental health institutions: Lentis, GGZ Friesland, GGZ Drenthe, Dimence, Mediant, GGNet Warnsveld, Yulius Dordrecht and Parnassia psycho-medical center The Hague. Maastricht: Maastricht University Medical Centre and the mental health institutions: GGZ Eindhoven en De Kempen, GGZ Breburg, GGZ Oost-Brabant, Vincent van Gogh voor Geestelijke Gezondheid, Mondriaan, Virenze riagg, Zuyderland GGZ, MET ggz, Universitair Centrum Sint-Jozef Kortenberg, CAPRI University of Antwerp, PC Ziekeren Sint-Truiden, PZ Sancta Maria Sint-Truiden, GGZ Overpelt, OPZ Rekem. Utrecht: University Medical Center Utrecht and the mental health institutions Altrecht, GGZ Centraal and Delta.

The research and publication of this thesis was financially supported by Research Institute SHARE (Science in Healthy Ageing and healthcaRE), University Medical Center Groningen (UMCG), University of Groningen, Department of Epidemiology, Rob Giel Research center (RGOc) and GROUP-project.

ISBN: 978-90-367-9882-2 (printed version)
ISBN: 978-90-367-9883-9 (electronic version)

Design: Rukhsana Rushmi
Cover: Loes kema, GVO drukkers & vormgevers B.V.
Lay-out: Md. Atiqul Islam
Printed by: GVO drukkers & vormgevers B.V., Ede, the Netherlands

university of groningen

Statistical approaches to explore clinical heterogeneity in psychosis

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. E. Sterken
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on
Wednesday 5 July 2017 at 09.00 hours

by

Md. Atiqul Islam
born on 1 January 1981
in Pabna, Bangladesh

Supervisor
Prof. E.R. van den Heuvel

Co-supervisors
Dr. R. Bruggeman
Dr. B.Z. Alizadeh

Assessment Committee
Prof. E.C. Wit
Prof. A.J. Oldehinkel
Prof. A.C. Krabbendam

To my parents, wife & daughter

Paranimfen

Sanne Y. Smith-Apeldoorn

Jan van Bebber

CONTENTS

CHAPTER 1

General Introduction

1. Introduction

1.1. Psychosis (schizophrenia spectrum disorder)

Schizophrenia and related psychotic disorders are now referred to as Schizophrenia Spectrum Disorders (DSM-5) or –by some- as Psychosis spectrum syndrome (van Os, 2016). Schizophrenia and related psychotic disorders consist of multiple symptom dimensions (van Os et al., 2010). A particular group of these non-affective psychotic disorders is a complex, multidimensional chronic brain disorder with a lifetime prevalence of nearly 1.5% (Perala et al., 2007; van Os et al., 2010).

The early onset of the disease, along with its chronic course, makes schizophrenia a debilitating disorder for many patients and their relatives (Mueser and Jeste, 2008). The other psychotic disorders, *i.e.* schizoaffective disorder, schizophreniform disorder, psychotic disorder not otherwise specified and brief psychotic disorder, cause similar symptoms of psychosis, albeit with a shorter duration of illness. Causes of schizophrenia are still to be elucidated, although there are some well-known risk factors. For example, individuals with a first-degree relative with schizophrenia have a higher risk of developing the disorder, about tenfold compared to general population (Gejman et al., 2010). There is evidence that the level of familial clustering of psychotic disorder is higher when people are living in urban environment or belong to a minority group (van Os et al., 2010). It is now recognized that high heritability (80%) estimates from classical twin and adoption studies is not only due to genetic influence, but also underlying intra-familial environmental effects that are moderated by gene–environment interactions (Gejman et al., 2010; van Os et al., 2010; van Os and Kapur, 2009).

Psychotic disorder is characterized by an array of heterogeneous symptoms including delusions, hallucinations, disorganized speech or behavior, and impaired cognitive ability occurring for a significant period of time during at least one month period and associated with continuous problems over at least a six-month period (Perala et al., 2007; van Os and Kapur, 2009). These multiple dimensions have also been mentioned in DSM-5 (American Psychiatric Association, 2013). The delusions (i.e. mostly false beliefs rooted in the mind based on incorrect inference) and the hallucinations (i.e. sensory-driven incidents that involve hearing or seeing something that is not reality based) are often considered the cardinal features of the illness of psychosis, partly because they are easy to identify and greatly affect functioning and society.

The signs and symptoms of schizophrenia may vary dramatically from person to person, both in pattern and severity. Recently, there has been a debate on the nature of the negative symptoms. Negative symptoms comprise in dysfunction of communication, affect and emotion, socialization, capacity of pleasure and motivation (Stahl and Buckley, 2007). According to authors such as Liemburg et al. (2013), one can discriminate two dimensions of negative symptoms; i.e. 1) social amotivation (social, emotional withdrawal and reflects diminished interest in or affective commitment to the social environment) and 2) expressive deficit (blunted affect, poverty of speech and motor retardation, and reflects diminished expressive responsiveness in verbal and non-verbal communication). These authors have clearly demonstrated the clinical validity of such dimensions of negative symptoms in first-episode patients with a psychotic disorder in cross-sectional studies (Liemburg et al., 2013). However, whether this two-dimension approach also holds for longitudinal studies needs to be demonstrated.

Cognitive impairment is another dimension of symptoms. There has been a resurgence of interest in the cognitive alterations of schizophrenia. It is often stated that patients with a diagnosis of schizophrenia have a broad-based cognitive impairment of, on average, about 1 SD below the norm across a range of cognitive abilities (attention, speed of processing, working and long-term memory, executive function, and social cognition) (Kahn RS, 2013; van Os et al., 2010; van Os and Kapur, 2009; Fioravanti et al., 2005). Furthermore, the uniformity of the cognitive impairments has been questioned. Quee et al. (2014) found both severe impairment and normal functioning in non-affected siblings and a mixed profile group in between these two. Interestingly, these cognitive profiles correlated well with their affected family-member, suggesting that cognition is diverse and heterogeneous in people with psychotic disorders (Quee et al., 2014).

In line with the more integral approach on symptom dimensions, physical complaints have increasingly been considered as the "somatic dimension" of schizophrenia spectrum disorders. Comorbidity with somatic disorders has now been recognized as an important factor, leading to a 15-20 years shorter life expectancy for patients with schizophrenia. These comorbid health-conditions may contribute up to 60 percent of the three times excess of premature mortality in schizophrenia (De Hert et al., 2011; Parks et al., 2006; Vreeland, 2007). Indeed, patients with schizophrenia have up to 54 percent metabolic syndrome (Bruins et al., 2016) and a 2-3 fold higher risk of diabetes mellitus (Bushe and Holt, 2004; van Winkel et al., 2006) and cardiovascular diseases (Bresee et al., 2010; De Hert et al., 2009; Hennekens et al., 2005).

Thus, psychiatric symptoms (e.g. on positive and negative symptoms) amended with cognitive impairments, and somatic comorbidity are usually heterogeneous and differ in origin, structure and clinical expression. Although this notion has been widely known for a long time (Markova and Berrios, 1995), these differences are often overlooked both clinically and statistically in research. Ignoring differences in structure between symptoms has also naturally yielded biased homogeneous structure of symptoms. With the growing awareness of the heterogeneity of psychotic disorders, there is also a growing need in classical and model-based statistical clustering approaches to clarify the underlying structures.

The main aim of the thesis is to explore the heterogeneity in cognitive functioning and clinical symptoms in schizophrenia patients and their unaffected siblings using cross-sectional and longitudinal data. This aim is achieved by applying statistical methods, such as classical clustering, linear mixed effects and group-based trajectory modeling techniques.

1.2. Statistical Analysis for Heterogeneity

1.2.1. Heterogeneity

To illustrate heterogeneity, consider for example clinical trials where some patients do not response to the treatment and others do respond well under the same treatment plan. Therefore, modest clinical effects can sometimes be misleading because they may be composed of a mixture of significant benefits for some, no benefits for many and harm for a few. The same would hold true for schizophrenia patients, where individual differences would mask general patterns.

In general, each subject would have its own profile. The average profile of all subjects would give the idea of having just one population profile. The average profile may provide valuable

information, assuming that the population under study is to large extent homogenous. For longitudinal studies, Verbeke and Lesare (1996) demonstrated that this homogeneous population is then described by a single mean trajectory and variance-covariance matrix. However, this assumption is highly unrealistic when subgroups of populations exist. In psychiatry, different disease symptoms or diagnostic groups could be classified by different mean profiles. Ignoring the heterogeneity can produce biased estimates of the association parameters and their corresponding variance terms (Verbeke and Lesaffre, 1996). To break down the "seemingly homogeneous population" into more meaningful subgroups with similar profiles or patterns, one may need to quantify the heterogeneity.

One way to dissect heterogeneity of a group of patients with seemingly the same diagnosis is to use the positive and/or negative symptoms or cognitive impairment scores and apply clustering techniques to form homogeneous symptom subtypes (Dawes et al., 2011; Jablensky, 2006; Joyce and Roiser, 2007). Cluster analysis is used to classify objects into groups (or clusters/classes/components) such that objects within a group are more similar than between groups. Forming clusters depend on study design and analytical tools. In cross-sectional studies, where the subjects are independent in measuring the outcome, clustering techniques, like hierarchical clustering and K-means clustering, can address the heterogeneity and form homogeneous subtypes. In longitudinal studies psychiatric symptoms (on positive and negative symptoms or cognitive impairment) of patients (or siblings) may be heterogeneous over time. They are likely to originate from latent distinct trajectory patterns. Thus it is important to capture the individual trajectories and to understand what affects them. Mixed effects models are the leading statistical techniques to describe individual trajectories. Moreover, these potentially existing heterogeneity in disease course patterns, warrant a method which identifies the presence of unobserved heterogeneity and to form homogeneous subgroups of patients. This leads to finite mixture modeling (Schlattmann, 2009) which is designed to identify clusters of individuals following similar pattern of progression of outcomes over time (Jones and Nagin, 2007). In other words, the primary goal is to identify groups of patients that have similar patterns of symptoms over time. Some methods for dealing with heterogeneity for cross-sectional and longitudinal data are briefly explained below.

1.2.2. Classical clustering

Hierarchical clustering is a classical clustering technique, which is often used in psychiatry. It produces a nested (i.e. hierarchical) sequence of clusters that can be presented in a dendrogram (Figure 1). Agglomerative and divisive hierarchical clustering are ways to form nested clusters. Agglomerative clustering is a bottom-up method. It starts with partitioning all sample objects to individual or separate clusters and then successively merging the closest pair of clusters using some kind of similarity criterion. It ends when all objects are in one cluster (Figure 1). The advantage is that it can produce an informative ordering of the objects and produce small clusters, which may be helpful for discovery. Divisive clustering is the opposite of agglomerative method; it starts with all objects in one cluster and then iteratively split clusters which are most dissimilar until all objects end up in their own cluster (top-down approach) (Figure 1). The disadvantage of hierarchical clustering is that if an object becomes a member of any cluster, it will neither be removed from that cluster nor

be mixed up with objects of any other clusters. It may cause incorrect groups at an early stage (Fernández and Gómez, 2008).

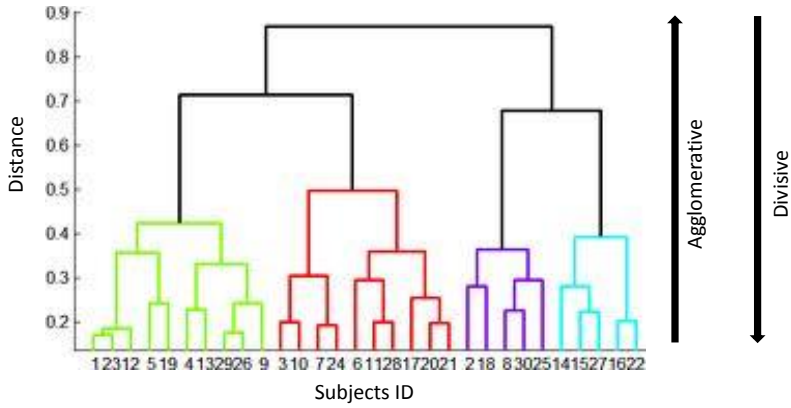


Figure 1: Dendrogram displaying number of clusters

In order to construct the clusters, there are different ways of measuring the distance between groups of objects such as single-linkage, complete-linkage, average-linkage and Ward's minimum variance method. The first three define the distance between clusters as the minimum, maximum and average distance between any two objects of the clusters. Ward's minimum variance method, on the other hand, finds the pair of clusters that leads to minimum increase in total within-cluster variance when merging two clusters as the distance measure (Ward, 1963).

Another popular set of clustering techniques are partitioning methods that attempts to divide the whole set of objects into a pre-defined number of clusters (Henry et al., 2005; Gordon, 1999; Borgen and Barnett, 1987; Hartigan and Wong, 1979). The most popular one is K-means clustering. It aims to minimize the sum of the squared distances between the objects and their cluster centers, by iteratively reallocating objects to the clusters until full convergence. To illustrate the procedure, let us assume that there are *k* clusters and choose *k* centroids in the data space of the objects (either randomly or by using hierarchical clustering). Then, assign each object to the closest centroid using a predetermined criterion (e.g. minimize the sum of the squared distances). When all objects are assigned to the *k* centroids, calculate the mean or median of the variables under study for the formed *k* groups. Then again assign each object to the new centroids and repeat the process until the groups do not change anymore.

Practically, clustering techniques provide little information about the cluster structure in the data. There is no unified approach on what essentially constitutes a cluster and no conclusive answer for choosing the number of clusters (Fernández and Gómez, 2008; Milligan and Cooper, 1985). Furthermore, another major difficulty is to estimate the threshold of the objective function (i.e. the within-cluster sum of squares) and the number of clusters when there is no information other than the observed values is available (Hartigan, 1975). This is true both for hierarchical and K-means clustering techniques. Since there is no *a priori* information on natural groupings or subtypes of patients or siblings on the basis of symptoms, I propose using hierarchical clustering to find the most appropriate number of clusters. The result of hierarchical clustering is a good option to obtain a priori

information and then use this solution as input for K-means clustering to finally form the subtypes of patients or siblings. However, one of the main problems in hierarchical clustering is to determine the true number of clusters since it produces an informative ordering of the objects and produce series of small clusters (Fernández and Gómez, 2008; Milligan and Cooper, 1985).

Over the past decades, several procedures/indices have been proposed for determining the number of clusters and testing the null hypothesis that there is no cluster structure in the dataset at all (Milligan and Cooper, 1985). The majority of existing indices does not test formally a null hypothesis but rather estimate a summary statistics that points towards an optimal number of clusters. These summary statistics are typically functions of the within clusters and the between clusters sum of squares. In addition, these indices are evaluated either locally or globally to determine the number of clusters with respect to the clustering algorithms. Global methods utilize entire dataset and maximize it as a function of the number of clusters. Most of the global methods are undefined for one cluster and hence there is no indication whether the data should be clustered at all. Local methods use individual pairs of clusters and test whether they would be merged or not (Tibshirani et al., 2001; Gordon, 1999). Several criteria or stopping rules for the indices have been provided in literature, but extensive simulation studies on realistic data were not conducted. Therefore, I will discuss these shortcoming and potential solutions in this dissertation.

1.2.3. Linear mixed models

Repeated measures on subjects are very common in health, social, behavioral and biological sciences. The major challenge in analyzing repeated measures data is the fact that the measurements on a subject are correlated. This correlation must be taken into account during the statistical analysis to obtain valid inference i.e. estimates of effect sizes for association between parameters and outcomes of interest. The correlation can often be captured by introducing random effects in the classical statistical analyses e.g. linear and logistic regression. These statistical models combine the components of fixed effects, random effects (e.g. random-intercept and random-slope), and repeated measurements in a single unified approach. These models are called linear mixed models (LMM) for continuous longitudinal outcomes (Laird and Ware, 1982; Verbeke and Molenberghs, 2000), and generalized linear mixed models (GLMM) for other type of outcomes (Breslow and Clayton, 1993; Molenberghs and Verbeke, 2006).

The mathematical equations for LMM are explained briefly here for longitudinal data and subject-specific time profiles. Let Y_{it} be the response measure (e.g. score of psychotic experiences) for subject i (=1, 2, …, *N*) measured at time T_{it}, t = 1, 2, …, n_i. The linear mixed model is simply defined in the matrix form as

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i b_i + \varepsilon_i$$

Where, $b_i \sim N(\mathbf{0}, D)$, $\varepsilon_i \sim N(\mathbf{0}, \Sigma_i)$, $\boldsymbol{Y}_i$ is the n_i-dimensional response vector for subject i, $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are the $(n_i \times p)$ and $(n_i \times q)$ design matrices of known covariates, $\boldsymbol{\beta}$ is the *p*-dimensional vector of population-average regression coefficients (fixed effects), b_i is the *q*-dimensional vector of random effects for subject i, ε_i is a n_i-dimensional vector of measurement error components. It is assumed that b_i and ε_i are independent. Conditional on the random effects b_i, the distribution of $\boldsymbol{Y}_i$ is given by $\boldsymbol{Y}_i | b_i \sim N(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i b_i, \Sigma_i)$. The inference is based on maximizing the likelihood function of the

marginal response $\boldsymbol{Y}_i$. More specifically, a subject-specific time trajectory of polynomial form can be defined as

$$Y_{it} = \beta_0 + \sum_{k=1}^{p}\sum_{r=0}^{q}(\beta_{rk}X_{ik} + b_{ri}) * T_{it}^{r} + \varepsilon_{it}$$

where, β_0 is the overall intercept, β_{rk} are the r^{th} (r = 0, 1, 2, …, q) polynomial form of time and k^{th} (r = 1, 2, …, p) fixed effects parameters, b_{0i} is the random-intercept, b_{ri} is the random-term for time order T_{it}^{r} of subject i, and ε_{it} is error disturbance term. The model parameters and variance components are estimated by either Maximum Likelihood (ML) or Restricted Maximum Likelihood (REML) estimation procedure (Verbeke and Molenberghs, 2000). Note that each subject has its own time profile in this model.

1.2.4. Group-based trajectory modeling

In literature, there are many approaches which are applied in many different situations to quantify homogeneous groups with their own shapes and patterns for longitudinal trajectories. To cluster subjects based on continuous longitudinal data, Verbeke and Lesaffre (1996) assumed a normal mixture in the distribution of random effects and applied their method to clustering of growth curves (Verbeke and Lesaffre, 1996). De la Cruz-Mesía et al (2008) proposed a mixture of nonlinear mixed models for describing nonlinear relationships across time and perform clustering of subjects on outcome (De la Cruz-Mesía et al., 2008). Other longitudinal approaches used to understand and group differences in developmental trajectories over time include latent class analysis which was primarily used for categorical or binary data (Vermunt and Magidson, 2003) or a semi-parametric mixture model that was appropriate for data with skewed distributions (Jones et al., 2001). The growth mixture modelling is one of the approaches when outcome variables are approximately normally distributed.

Group-based trajectory modeling (GBTM) is a semi-parametric statistical method for analyzing developmental trajectories, which means describing the evolution of an outcome over time (Nagin, 1999). GBTM is sometimes called latent class growth modeling (Andruff et al., 2009). GBTM is an application of finite mixture modeling and is designed to identify clusters of individuals following similar patterns of change of some behavioral, biological, physical outcome (e.g. cognition, negative symptom) over time (Jones and Nagin, 2007; Nagin, 2014). Traditional growth curve modeling techniques assume that subjects come from a single population and estimate a single trajectory that averages the individual trajectories of all subjects in a given sample. This average trajectory comprises the averaged intercept and slope for the entire sample. This approach captures individual differences by estimating random coefficients that represents the variability in the intercept and slope. But GBTM fixes the slope and intercept for the subgroup of individuals having a similar trajectory, given that individual differences are captured by the multiple trajectories included in the model (Andruff et al., 2009). GBTM is also a flexible statistical tool for identifying and summarizing the homogenous group of individuals and observing their development patterns over time in the form of both graphical and tabular way, which makes it easier to understand. Another important feature is that the GBTM also takes into account drop-out of participants over time (Haviland et al., 2011).

Mathematically, let Y_{it} be the longitudinal sequence of measurements (e.g. cognition or subdomains of negative symptoms) on individual *i* (= 1, 2, …, *N*) over time *t* (= 1, 2, …, *T*). The GBTM assumes that the population is composed of a mixture of *g* (=1, 2, …, *G*) underlying trajectory groups with marginal density $f(y_i) = \sum_g \pi_g p^g(y_i)$, where $p^g(y_i)$ is the density function of Y_i given its membership in group *g* and π_g is the probability of belonging to group *g*. The basic GBTM assumes that the random variables, Y_{it} are independent given the condition on membership in group *g*, therefore $p^g(y_i) = \prod_{t=1}^{T} p^{gt}(y_{it})$. The group membership probabilities, π_g are estimated by a multinomial *logit* function as $\pi_g = \exp(\theta_g)/\sum_{g=1}^{G} \exp(\theta_g)$, where θ_1 is standardized to zero so that estimation of each probability of π_g stays between 0 and 1.

In this thesis, I will apply the censored normal model (CNORM). The CNORM model is useful for modeling the conditional distribution of psychometric scale data, given group membership (Jones et al., 2001; Nagin and Tremblay, 2001). A normal distribution allowing for censoring is used because the data tend to cluster at the minimum (Min) and at the maximum (Max) of the scale. In our case, $p^{gt}(y_{it})$ is assumed to follow the censored normal distribution to accommodate the possibility of clustering at the value minimum and maximum. The likelihood of observing the data trajectory for individual *i*, given he/she belongs to group *g*, is given by

$$p^g(y_i) = \prod_{y_{it}=Min} \Phi\big((Min - \mu_{it}^g)/\sigma\big) \prod_{Min<y_{it}<Max} \frac{1}{\sigma}\varphi\big((y_{it} - \mu_{it}^g)/\sigma\big) \prod_{y_{it}=Max} \Big(1 - \Phi\big((Min - \mu_{it}^g)/\sigma\big)\Big)$$

Where, $\mu_{it}^g = \beta_0^g + \beta_1^g time_{it} + \beta_2^g time_{it}^2 + \beta_3^g time_{it}^3$ be the mean group time profile for the symptom/cognitive measurement in group *g* (Jones et al., 2001). Likewise growth curve modeling, a polynomial relationship is used to model the link between period and cognition/symptoms of individuals. The model assumes up to third-order polynomial relationship between μ_{it}^g and period (e.g. follow-up time) (Jones and Nagin, 2007).

1.3. Statistical Analysis for Associations

1.3.1. Modeling association

Many studies have shown a general pattern that patients are being more affected than normal controls with respect to reporting problems on functioning and symptoms on a population level. In majority of these studies non-affected siblings display values somewhere between these two groups (Quee et al., 2014; Krabbendam et al., 2005). To be able to compare groups of subject that are independent, Pearson chi-square (for categorical outcome) or ANOVA (for continuous outcome) is applicable. However, when subjects belong to the same family (as in patient-sibling studies), the group comparisons may be done using linear or generalized linear mixed effects models taking into account the familial correlation. This may lead to confirmation of familial liability, meaning that patients having more diseases than their unaffected siblings and healthy controls. One may also use the intra-cluster correlation coefficient (ICC) to calculate the measure of the relatedness/correlations (e.g. familial liability) of the subjects *within* the family.

1.3.2. Mixture distribution modeling

As mentioned earlier, siblings are at risk to develop psychotic experiences. A limited number of them make the transition to a clinical-defined psychosis over time. Psychotic experiences are also heterogeneous over time and this heterogeneity can possibly be explained by different factors. To learn more about the development of psychotic experiences over time, one can study factors that are known to predict psychosis. In this case the outcome is an increase or decrease of the number of psychotic experiences, measured on a numerical scale. This outcome is most likely not normally distributed, since a large number of siblings may have no psychotic experiences at all while a small number of siblings do actually report psychotic experiences. In other words, the distribution of the outcome is at least highly skewed to right. However, given the heterogeneous nature of the reported psychotic experiences, the distribution is more likely to be a bi-modal distribution (Figure 2). Such a distribution can be described by a mixture distribution of known parametric (like the normal) distributions.

For a unimodal distribution, the classical approach, for example LMM or GLMM, can be used to estimate the effect of the predictors on the outcome. This would still be the case even when the distribution of outcome is highly skewed. However, when there are concerns with a lot of zeros in the observed continuous outcome, the distribution of the outcome would display at least a bi-modal distribution. Classical statistical approaches may not lead to the correct estimates of the effects of the predictors. An alternative approach, which can deal with the excessive number of zeros, is then preferred to obtain unbiased estimates. The group of subjects with no psychotic experiences can be scored *zero*, while the group with experiences is labeled *one.* These two groups can be described with a Bernoulli distribution (scoring zero or one). Then the next step is to consider those subjects, who report experiences (i.e. nonzero), and describe their continuous outcome with a skewed distribution, like the lognormal distribution. The final step is then to combine these two models and to estimate the parameters for the probability of having a nonzero value (e.g. using generalized linear mixed effects models) together with the parameters for the skewed distribution (e.g. using a random effects lognormal model). The random effects in the binary and continuous parts are needed to address the repeated measurements over time (Tooze et al., 2002; Olsen and Schafer, 2001; Smith et al., 2015). This type of modeling is referred to as random effects mixture modeling and it enables to draw conclusions on the factors that may associate the binary part and/or the continuous part.

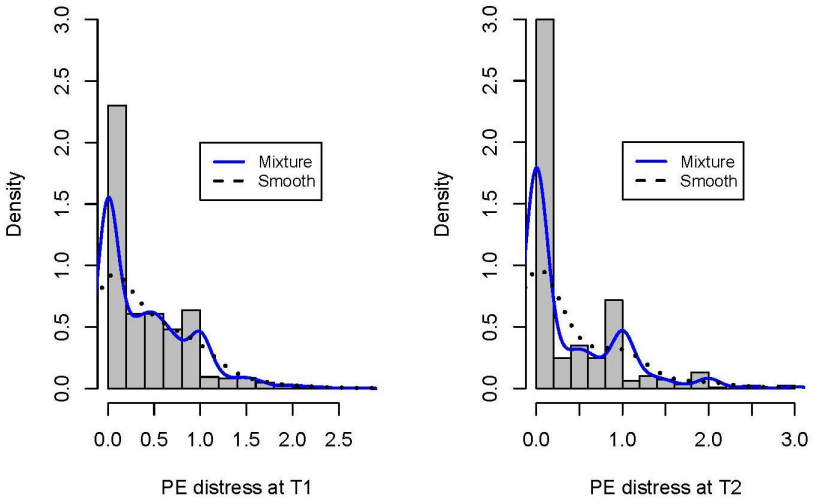


Figure 2: Empirical density, mixture density and smooth density of PE distress at time T1 and T2.

1.3.3. Dealing with missingness

Missing data are inevitable in any study, but in particular when data are collected repeatedly over time. The collection of a complete dataset on subjects is almost impossible. Depending on the nature of the study, missingness will appear in various forms. For instance, in a cross-sectional survey, missing data is usually of the form of item non-response. In this case a subject is not able or does not want to respond to a particular question or measurement. In longitudinal studies, attrition is the most common missing data problem. Here subjects drop out of the study prematurely before its termination and do not return. The pattern of attrition is an example of missing data for which the incompleteness can be ordered monotonically. Attrition, also referred to as dropout, may not be the only form of missing data in a longitudinal setting. For instance, a subject can miss several observation periods but eventually returns to the study. The latter type of missingness is often referred to as intermittent missingness.

To handle missing data, the mechanisms that lead to these missing values are of main importance. For instance, what drives the missingness patterns? To be more specific, is there a relationship between the missing data and the underlying values in the dataset. Three types of missing-data mechanisms have been defined in the literature. A process is said to be missing completely at random (MCAR) if the missingness is independent of both the unobserved and observed data. This implies that the probability of missing data on response is unrelated to the value of response itself or to the values of any other variables in the dataset. If the missingness depends on the observed data but it is independent of the unobserved values then the missing mechanism is said to be missing at random (MAR). MAR implies that the probability of having missing data is unrelated to the values that were missing, conditional on the observed values. If the process is neither MCAR nor MAR it is missing not at random (MNAR). The process then depends on the unobserved measurements. The assumption of MNAR does imply that the probability of a measurement being missing depends on the unobserved data (Rubin, 1976; Verbeke and Molenberghs, 2000; Little and Rubin, 2002; Allison, 2002).

Various simple methods, such as complete case analysis (CC) and last observation carried forward (LOCF), for handling missing data are available. Generally, simple methods such as complete case analysis (CC) work under the assumption that the missing mechanism is MCAR and in some special cases of MAR. A complete case analysis includes all subjects that would have all data recorded (relevant to the analysis). However, the method suffers from severe drawbacks: loss of valuable information, biased estimates, and inefficient estimates. The method of last observation carried forward (LOCF) replaces every missing value by the last observed value from the same subject. This method can be used both for monotone and non-monotone missing data but it is typically used in situations where incompleteness is due to attrition. Like other single imputation methods it overestimates the precision by treating imputed values and observed values on equal footing (Molenberghs and Verbeke, 2006; Beunckens et al., 2005). More importantly though even under the very strong assumption of MCAR, LOCF can be biased.

Multiple Imputation (MI) is currently one the most popular methods to deal with missingness under the MAR assumption. The idea of multiple imputation procedure is to replace each missing observation in the dataset with M plausible values, creating a set of M fully complete data sets. To analyze the data, each imputed data set is analyzed separately using conventional analysis method and programs. The results are then pooled in such a manner that the uncertainty in the imputed values averages out and disappears (Verbeke and Molenberghs, 2000; Little and Rubin, 2002; van Buuren, 2007). Maximum likelihood (ML) method sometimes provides valid inferences under MAR assumption. The approach gives appropriate estimates when the missingness occurs only in the (repeated) outcomes. If risk factors or predictors are missing ML may lead to biased estimates. Pattern mixture modeling, on the other hand, deals with missingness under the MNAR assumption (Verbeke and Molenberghs, 2000). It studies the statistical model conditional on the missing data indicators. ML, MI or Bayesian MI may provide biased estimates under MNAR (Schafer and Graham, 2002).

1.4. Database and software

To determine the aims of the thesis, data were obtained from the Genetic Risk and Outcome of Psychosis (GROUP) project, a longitudinal multicenter cohort study in the Netherlands and Belgium from April 2004 t/m in December 2013. The GROUP project provides a rich cohort data set on patients with schizophrenia, their unaffected siblings, and healthy controls at baseline, three and six years of follow-up. Patients were identified from a representative set of clinicians by screening their caseload and evaluating the patients to the inclusion criteria. Subsequently, a group of patients presenting consecutively at these services either as out-patients or in-patients were recruited for the study. In order to test hypotheses about the aetiology of non-affective psychosis, a cohort of family members e.g. siblings and parents with resilience for psychosis was being included. Controls were selected through a system of random mailings to addresses in the catchment areas of the cases. GROUP study examines vulnerability factors and protective factors for developing a psychotic disorder and the course thereof (Korver et al., 2012).

Statistical software such as Statistical Analysis System (SAS) version 9.4 and RStudio version 0.97.551/0.99.902 (R version 3.0.1/3.1.1) were used to perform all analyses throughout the thesis.

1.5. Outline of the thesis

The general aim of the thesis is to explore the heterogeneity in cognitive functioning and clinical symptoms in schizophrenia patients and their unaffected siblings using cross-sectional and longitudinal data. To this end, this thesis decomposes mainly two parts of statistical modeling. Part A contains statistical analysis for heterogeneity, here I will apply and evaluate classical clustering technique, linear and generalized linear mixed effects modeling, and group-based trajectory modeling. Part B includes statistical modeling for associations, here I will apply classical ordinal logistic regression, Cox-regression, generalized linear mixed models and mixtures of generalized linear mixed effects modeling.

From clinical perspectives, the current thesis describes a number of studies focusing on cognition, clinical symptoms, functional outcomes and disease outcomes in patients with psychosis, their unaffected siblings and healthy controls. Cognitive and symptoms heterogeneity have been characterized in this thesis.

In **chapter 2**, I investigate fourteen cluster indices to identify the correct number of subtypes for cross-sectional data. I use simulations that were based on a real case study with eight cognitive measures. I compared the indices on their performances for hierarchical clustering of subjects, while the simulations generated mixture distributions of multivariate normal distributions. I will show how well indices predict the simulated pre-defined number of clusters and also determine whether they can be used to decide if there would exist multiple subtypes at all.

In **chapter 3**, I apply the GBTM to identify homogeneous cognitive trajectories of patients with schizophrenia and their unaffected siblings over time. Distinct trajectories of composite cognitive functioning over time are identified and they distinguish different trajectories between patients and siblings, respectively. After finding these meaningful homogeneous trajectories of patients and siblings, I examine whether patients' profiles predict cognitive profiles of their unaffected sibling.

Chapter 4 also describes the application of GBTM to determine homogenous groups of patients based on symptoms (social amotivation and expressive deficits of negative symptom sub-domains) over time. The aim is to determine if these homogeneous groups contribute to the understanding of subdomains of negative symptoms and investigate if their specific trajectories impact functioning and quality of life. Additionally, the application of multiple imputation techniques are used to deal with missing data on outcomes and other covariates. Next, I move to the association of heterogeneous *outcomes* and candidate risk factors.

In **chapter 5**, the predictive value of neurocognitive and social cognitive measures on the course and impact of psychotic experiences in siblings of people with psychotic disorders is investigated using mixture distribution model. The methodology of mixture of generalized linear mixed effects modeling is also explained in this chapter. The application of MI techniques is also used here.

In **chapter 6**, I describe the heterogeneity, regarding somatic diseases and complaints among patients with psychotic disorders, their unaffected siblings and healthy population. I examine the effects of gender, age and familial liability on the prevalence of multimorbidity.

In **chapter 7**, I investigate the factors that contribute to DUP in a large sample that represents the treated prevalence of non-affective psychotic disorders. DUP is categorized into meaningful ordinal groups and the ordinal logistic regression is applied to identify important factors. Other statistical approaches are also used to confirm factors associated with DUP and discuss as well.

Chapter 8 finally synthesizes the main findings and discusses these from multiple perspectives: from methodological consideration and clinical implication. I conclude with some future perspective and suggestions for further research.

References

Allison, P.D. (2002). *Missing Data*. Thousand Oaks, CA: Sage Publications.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Washington, DC: American Psychiatric Publishing.

Andruff, H., Carraro, N., Thompson, A., Gaudreau, P., Louvet, B. (2009). Latent Class Growth Modelling: A Tutorial. *Tutorials in Quantitative Methods for Psychology* 5:11-24.

Beunckens, C., Molenberghs, G., Kenward, M.G. (2005). Direct likelihood analysis versus simple forms of imputation for missing data in randomized clinical trials. *Clinical trials (London, England)* 2:379-386.

Borgen, F.H., Barnett, D.C. (1987). Applying cluster-analysis in counseling psychology research. *Journal of counseling psychology* 34:456-468.

Bresee, L.C., Majumdar, S.R., Patten, S.B., Johnson, J.A. (2010). Prevalence of cardiovascular risk factors and disease in people with schizophrenia: a population-based study. *Schizophrenia research* 117:75-82.

Breslow, N.E., Clayton, D.G. (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association* 88:9-25.

Bruins, J., Pijnenborg, M.G., Bartels-Velthuis, A.A., Visser, E., van den Heuvel, E.R., Bruggeman, R., Jorg, F. (2016). Cannabis use in people with severe mental illness: The association with physical and mental health--a cohort study. A Pharmacotherapy Monitoring and Outcome Survey study. *Journal of psychopharmacology (Oxford, England)* 30:354-362.

Bushe, C., Holt, R. (2004). Prevalence of diabetes and impaired glucose tolerance in patients with schizophrenia. *The British journal of psychiatry.Supplement* 47:S67-71.

Dawes, S.E., Jeste, D.V., Palmer, B.W. (2011). Cognitive profiles in persons with chronic schizophrenia. *Journal of clinical and experimental neuropsychology* 33:929-936.

De Hert, M., Correll, C.U., Bobes, J., Cetkovich-Bakmas, M., Cohen, D., Asai, I., Detraux, J., Gautam, S., Moller, H.J., Ndetei, D.M., Newcomer, J.W., Uwakwe, R., Leucht, S. (2011). Physical illness in patients with severe mental disorders. I. Prevalence, impact of medications and disparities in health care. *World psychiatry : official journal of the World Psychiatric Association (WPA)* 10:52-77.

De Hert, M., Dekker, J.M., Wood, D., Kahl, K.G., Holt, R.I., Moller, H.J. (2009). Cardiovascular disease and diabetes in people with severe mental illness position statement from the European Psychiatric Association (EPA), supported by the European Association for the Study of Diabetes (EASD) and the European Society of Cardiology (ESC). *European psychiatry : the journal of the Association of European Psychiatrists* 24:412-424.

De la Cruz-Mesía, R., Quintana, F.A., Marshall, G. (2008). Model-based clustering for longitudinal data. *Computational Statistics & Data Analysis* 52:1441-1457.

Fernández, A., Gómez, S. (2008). Solving Non-uniqueness in Agglomerative Hierarchical Clustering Using Multidendrograms. *Journal of Classification* 25:43-65.

Fioravanti, M., Carlone, O., Vitale, B., Cinti, M.E., Clare, L. (2005). A meta-analysis of cognitive deficits in adults with a diagnosis of schizophrenia. *Neuropsychology review* 15:73-95.

Gejman, P.V., Sanders, A.R., Duan, J. (2010). The Role of Genetics in the Etiology of Schizophrenia. *The Psychiatric clinics of North America* 33:35-66.

Gordon, A.D. (1999). *Classification*. London, UK: Chapman & Hall/CRC Monographs on Statistics & Applied Probability.

Hartigan, J.A. (1975). *Clustering Algorithms. A Wiley Publication in Applied Statistics, John Wiley & Sons, New York.*

Hartigan, J.A., Wong, M.A. (1979). A K-Means Clustering Algorithm. *Applied Statistics* 28:100-108.

Haviland, A.M., Jones, B.L., Nagin, D.S. (2011). Group-based Trajectory Modeling Extended to Account for Nonrandom Participant Attrition. *Sociological Methods & Research* 40:367-390.

Hennekens, C.H., Hennekens, A.R., Hollar, D., Casey, D.E. (2005). Schizophrenia and increased risks of cardiovascular disease. *American Heart Journal* 150:1115-1121.

Henry, D.B., Tolan, P.H., Gorman-Smith, D. (2005). Cluster analysis in family psychology research. *Journal of family psychology : JFP : journal of the Division of Family Psychology of the American Psychological Association (Division 43)* 19:121-132.

Jablensky, A. (2006). Subtyping schizophrenia: implications for genetic research. *Molecular psychiatry* 11:815-836.

Jones, B.L., Nagin, D.S., Roeder, K. (2001). A SAS Procedure Based on Mixture Models for Estimating Developmental Trajectories. *Sociological Methods & Research* 29:374-393.

Jones, B.L., Nagin, D.S. (2007). Advances in Group-Based Trajectory Modeling and an SAS Procedure for Estimating Them. *Sociological Methods & Research* 35:542-571.

Joyce, E.M., Roiser, J.P. (2007). Cognitive heterogeneity in schizophrenia. *Current opinion in psychiatry* 20:268-272.

Kahn RS, K.R. (2013). Schizophrenia is a cognitive illness: Time for a change in focus. *JAMA Psychiatry* 70:1107-1112.

Korver, N., Quee, P.J., Boos, H.B.M., Simons, C.J.P., de Haan, L., GPOUPinvestigators, Investigators, G. (2012). Genetic Risk and Outcome of Psychosis (GROUP), a multi-site longitudinal cohort study focused on gene-environment interaction: objectives, sample characteristics, recruitment and assessment methods. *International Journal of Methods in Psychiatric Research* 21:205-221.

Krabbendam, L., Myin-Germeys, I., Hanssen, M., van Os, J. (2005). Familial covariation of the subclinical psychosis phenotype and verbal fluency in the general population. *Schizophrenia research* 74:37-41.

Laird, N.M., Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics* 38:963-974.

Lieburg, E., Castelein, S., Stewart, R., van, d.G., Aleman, A., Knegtering, H. (2013). Two subdomains of negative symptoms in psychotic disorders: established and confirmed in two large cohorts. *Journal of psychiatric research* 47:718-25.

Little, R.J.A., Rubin, D.B. (2002). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.

Markova, I.S., Berrios, G.E. (1995). Mental symptoms: are they similar phenomena? The problem of symptom heterogeneity. *Psychopathology* 28:147-157.

Milligan, G.W., Cooper, M.C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika* 50:159-179.

Molenberghs, G., Verbeke, G. (2006). *Models for Discrete Longitudinal Data*. New York: Springer-Verlag.

Mueser, K. T. and Jeste, D. V. (2008). Lavretsky, H: History of Schizophrenia as a Psychiatric Disorder
. In *Clinical Handbook of Schizophrenia.* 3-12. New York: Guilford Press.

Nagin, D.S., Tremblay, R.E. (2001). Parental and early childhood predictors of persistent physical aggression in boys from kindergarten to high school. *Archives of General Psychiatry* 58:389-394.

Nagin, D.S. (1999). Analyzing Developmental Trajectories: A Semiparametric, Group-Based Approach. *Psychological Methods* 4:139-157.

Nagin, D.S. (2014). Group-based trajectory modeling: an overview. *Annals of Nutrition & Metabolism* 65:205-210.

Olsen, M.K., Schafer, J.L. (2001). A Two-Part Random-Effects Model for Semicontinuous Longitudinal Data. *Journal of the American Statistical Association* 96:730-745.

Parks, J., Svendsen, D., Singer, P., and Foti, M. (eds) (2006). *Morbidity and mortality in people with serious mental illness*. Alexandria: Alexandria: National Association of State Mental Health Program Directors (NASMHPD) Medical Directors Council.

Perala, J., Suvisaari, J., Saarni, S.I., Kuoppasalmi, K., Isometsa, E., Pirkola, S., Partonen, T., Tuulio-Henriksson, A., Hintikka, J., Kieseppa, T., Harkanen, T., Koskinen, S., Lonnqvist, J. (2007). Lifetime prevalence of psychotic and bipolar I disorders in a general population. *Archives of General Psychiatry* 64:19-28.

Quee, P.J., Alizadeh, B.Z., Aleman, A., van den Heuvel, E.R., GROUP Investigators. (2014). Cognitive subtypes in non-affected siblings of schizophrenia patients: characteristics and profile congruency with affected family members. *Psychological medicine* 44:395-405.

Rubin, D.B. (1976). Inference and missing data. *Biometrika* 63:581-592.

Schafer, J.L., Graham, J.W. (2002). Missing data: our view of the state of the art. *Psychological methods* 7:147-177.

Schlattmann, P. (2009). *Medical Applications of Finite Mixture models*. Berlin, Germany: Springer.

Smith, V.A., Neelon, B., Preisser, J.S., Maciejewski, M.L. (2015). A marginalized two-part model for longitudinal semicontinuous data. *Statistical methods in medical research*1-24.

Stahl, S.M., Buckley, P.F. (2007). Negative symptoms of schizophrenia: a problem that will not go away. *Acta Psychiatrica Scandinavica* 115:4-11.

Tibshirani, R., Walther, G., Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society.Series B, Statistical methodology* 63:411-423.

Tooze, J.A., Grunwald, G.K., Jones, R.H. (2002). Analysis of repeated measures data with clumping at zero. *Statistical methods in medical research* 11:341-355.

van Buuren, S. (2007). Multiple Imputation of Discrete and Continuous Data by Fully Conditional Specification. *Statistical methods in medical research* 16:219-242.

van Os, J., Kapur, S. (2009). Schizophrenia. *Lancet (London, England)* 374:635-645.

van Os, J., Kenis, G., Rutten, B.P. (2010). The environment and schizophrenia. *Nature* 468:203-212.

van Os, J. (2016). "Schizophrenia" does not exist. *BMJ* 352:.

van Winkel, R., De Hert, M., Van Eyck, D., Hanssens, L., Wampers, M., Scheen, A., Peuskens, J. (2006). Screening for diabetes and other metabolic abnormalities in patients with schizophrenia and schizoaffective disorder: evaluation of incidence and screening methods. *The Journal of clinical psychiatry* 67:1493-1500.

Verbeke, G., Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer series in statistics.

Verbeke, G., Lesaffre, E. (1996). A Linear Mixed-Effects Model With Heterogeneity in the Random-Effects Population. *Journal of the American Statistical Association* 91:217-221.

Vermunt, J.K., Magidson, J. (2003). Latent class models for classification. *Computational Statistics & Data Analysis* 41:531-537.

Vreeland, B. (2007). Treatment decisions in major mental illness: weighing the outcomes. *The Journal of clinical psychiatry* 68 Suppl 12:5-11.

Ward, J.H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 58:236.

Part A: Statistical Analysis for Heterogeneity

CHAPTER 2

A comparison of indices for identifying the number of clusters in hierarchical clustering: A study on cognition in schizophrenia patients

Md. Atiqul Islam

Behrooz Z. Alizadeh

Edwin R. van den Heuvel

GROUP Investigators

Communications in Statistics: Case Studies, Data Analysis and Applications, 1(2): 98-113

Abstract

Finding clusters in a complex dataset is not straightforward. Different indices were developed to quantify the number of clusters. Their performances were studied using unrealistic simulations, since they were considered at low dimensions. We investigated fourteen indices for eight dimensional data using simulations based on cognition measures. We focused on hierarchical clustering with Ward's agglomerative technique. Results indicated that Duda and Hart, Hartigan and Gap/pc were best performing. They estimated the number of clusters within ±1 with high probabilities. Duda and Hart index was most consistent while Gap/pc and WGap/pc together made a good distinction between single and multiple clusters.

Keywords: cluster indices; cluster analysis; hierarchical clustering; homogeneous subgroups; number of clusters

1. Introduction

1.1. Background

Clustering of data is used to classify groups of objects that are similar to one another within groups but different from each other between groups according to a defined mathematical criterion (e.g. Euclidean distance). In psychiatry, clustering is frequently used to group patients (e.g. suffering from schizophrenia) or their siblings to form homogeneous subtypes on the basis of variables or characteristics (e.g. cognition) that are related to the disease. Homogeneous subtypes may help explain or identify the disease severity (Cornblatt and Keilp, 1994; Chen and Faraone, 2000; Joyce and Roiser, 2007; Lin et al., 2009), in particular when the formed subtypes can be related to other clinical outcomes. In psychology, clustering is commonly used in family research to form family subtypes based on criteria such as parenting practices, church attendance, youth self-esteem, ethnic identity, personality, and patterns of change in marriage (Bray et al., 1995; Mandara, 2003; Henry et al., 2005).

Although there are highly complex clustering methods, common clustering techniques are hierarchical methods, which produce a nested sequence of clusters, and partitioning methods (e.g. K-means), which divide the whole set of objects into a pre-defined number of clusters (Hartigan and Wong, 1979; Borgen and Barnett, 1987; Gordon, 1999; Henry et al., 2005; Silver and Shmoish, 2008). Since partitioning methods are sensitive to the starting point (initial centroids) of formulating clusters, hierarchical clustering is typically used to obtain these initial centroids. Good reviews on these methods have been determined elsewhere (Clatworthy et al., 2005; Everitt et al., 2011).

In practice, a priori information about the actual groupings of subjects is typically missing, so it is necessary to identify the number of clusters on the basis of the observed data itself. Since hierarchical clustering is often used as initial step in formulating clusters, this technique can be helpful in defining the most likely number of clusters in combination with additional measures or criteria. Unfortunately, quantifying the true number of clusters has been a serious challenge in hierarchical clustering, since it produces an informative ordering of the objects and produce series of small clusters (Milligan and Cooper, 1985; Fernández and Gómez, 2008). Over the past decades, various measures or indices have been proposed for determining the number of clusters (Beale, 1969; Duda and Hart, 1973; Calinski and Harabasz, 1974; Hartigan, 1975; Sarle, 1983; Krzanowski and Lai, 1988; Kaufman and Rousseeuw, 1990; Tibshirani et al., 2001; Yan and Ye, 2007; Boone, 2011; Albalate et al., 2011; Albatineh and Niewiadomska-Bugaj, 2011), albeit not all indices originated from hierarchical clustering.

Milligan and Cooper (1985) applied the hierarchical clustering technique with a comprehensive Monte Carlo simulation to study thirty different procedures for determining the number of clusters. They suggested that the Calinski and Harabasz (CH) index (Calinski and Harabasz, 1974) and the Duda and Hart (DH) index (Duda and Hart, 1973) were the best performers followed by the C index (Dalrymple-Alford, 1970; Hubert and Levin, 1976), Beale (B) index (Beale, 1969), Cubic Clustering Criteria (CCC) (Sarle, 1983) and Hartigan (H) index (Hartigan, 1975). Despite the large number of indices already available in 1985, other indices have been developed since. Krzanowski and Lai (1988) modified the Marriott index (Marriott, 1971) using the within-group sum of squares as the objective function rather than the within-group determinant and demonstrated that the KL index

was superior to the Marriott index based on simulation studies. Kaufman and Rousseeuw (1990) introduced the *silhouette* (S) index which is based on the comparison of its "tightness" and "separation" using a dissimilarity matrix (e.g. Euclidean distance). They concluded that the highest average *silhouette* width determines the true number of clusters. Tibshirani et al. (2001) developed two Gap statistics (Gap/uniform and Gap/principal components) and compared them with the KL and the S index for K-means clustering. They concluded that the Gap/pc was better than the other indices. Yan and Ye (2007) proposed the weighted Gap (WGap) and the difference of difference-weighted Gap (DD-WGap) statistics using the weighted within-clusters sum of errors (a measure of the within-clusters homogeneity). They used K-means clustering and compared these two methods with the original Gap (uni/pc) statistics in simulation studies. They indicated that the WGap and the DD-WGap statistics were highly effective in determining the number of clusters.

Since the indices B, DH, CH, H and KL are all functions of the same within sums of squares, these indices can in principle be considered identical. However, they all used different functions or calculations with their own developed criteria to identify the true number of clusters. These criteria were based on different distributional characteristics. This makes it hard to compare them mathematically, thus we resorted to simulations. Furthermore, the vast literature on these indices also compared them with simulations, but mostly to relatively simple and nonrealistic settings, e.g. the dimension of the clustering variables were restricted to small numbers. In practice, the dimensions are typically larger (e.g. six to eight variables) and it is unclear if all variables truly contribute to the formation of homogeneous subtypes. This paper evaluates 14 indices for hierarchical clustering by simulation, implementing these practical settings on the basis of a real case study. We studied B, DH, CH, H, C-index, CCC, KL, S, Gap, WGap, and DD-WGap indices since these indices have been reported in literature as good performers. Some of the indices in this study are based on the (within and between clusters) sum of squares (B, DH, CH, H, CCC, KL, and Gap) and some others are calculated from a dissimilarity matrix (C and S). We will investigate how well they predict the simulated number of clusters, but we will also determine whether they can be used to decide if there would exist multiple clusters or not. Details on these indices are discussed in section 2, while the motivating example, which is used as input for the simulation study, is provided next. Details on the simulation study are provided in section 3 followed by the results of the simulation studies in section 4. Finally, Section 5 provides conclusions and final comments.

1.2. Motivating Example

Cognitive heterogeneity is an obstacle to clarify the neuropathological foundations of schizophrenia. Heterogeneity in cognition may possibly be addressed adequately using clustering techniques to form homogeneous cognitive subtypes (Jablensky, 2006; Joyce and Roiser, 2007; Dawes et al., 2011) to identify disease (schizophrenia) severity. In our study, data was extracted from the Genetic Risk and Outcome of Psychosis (GROUP), a longitudinal multi-center cohort study in the Netherlands and Belgium. A detailed description of the GROUP study has been published elsewhere (Korver et al., 2012). In brief, it comprised measures on cognitive functioning, clinical symptoms and genetic make-up of patients, their unaffected siblings and parents, and unrelated controls at multiple time points. Eight neurocognitive measures were obtained: Continuous Performance Test-HQ (CPT-HQ) and

Standard deviation of CPT-HQ (attention/vigilance) (Stinissen et al., 1970; Quee et al., 2011), Word Learning Task (WLT) Immediate and Delayed Recall (verbal learning and memory) (Brand and Jolles, 1985), Wechsler Adult Intelligence Scale-III (WAIS-III) Digit Symbol Coding (processing speed), WAIS-III Arithmetic (working memory), WAIS-III Block Design (reasoning and problem solving), and WAIS-III Information (verbal comprehension) (Blyler et al., 2000). These cognitive traits have been described in detail elsewhere (Meijer et al., 2012; Quee et al., 2014). We obtained data on 860 independent patients with 77.4% male and mean age 27.22 years (SD=7.46) and 439 independent controls with 48.1% male and mean age 30.46 years (SD=10.53) with complete cognitive measures and all from different families at baseline. The mean ± standard deviation of the raw scores of all cognitive variables were as follows: CPT performance index (220.95 ± 63.05), CPT standard deviation (92.90 ± 36.29), Block Design (40.28 ± 17.02), Digit Symbol (65.01 ± 15.98), Arithmetic (12.17 ± 4.79), Information (16.66 ± 5.51), WLT Immediate Recall (22.91 ± 6.11), and Delayed Recall (7.52 ± 2.85). The controls were used to obtain z-scores on patients for all eight cognitive measures that were age and gender specific, and these z-scores were used for clustering.

The number of clusters suggested by the fourteen selected indices using hierarchical (Ward's agglomerative) clustering are provided in Table 1. The dendrogram (Supplementary Figure 1) is presented in the Supplementary file. The indices do not agree on determining the number of clusters. The B index suggests that the patients represent one homogeneous group, while the CH, H, S, CCC, Gap/uni, Gap/pc, WGap/uni, and WGap/pc all suggest just two cognitive subtypes. Four subtypes are indicated by the KL, DD-WGap/uni and DD-WGap/pc, while the C-index indicates one subtype fewer and the DH even indicates five subtypes. Based on these results, it would be difficult to determine the correct number of clusters, unless we know which indices are most reliable for this type of data.

Table 1: The number of clusters determined by different indices from GROUP data

Number of clusters	Indices
1	B
2	CH, H, CCC, S, Gap/uni, Gap/pc, WGap/uni, WGap/pc
3	C
4	KL, DD-WGap/uni, DD-WGap/pc
5	DH

2. Methods

2.1. Overview of the selected indices

Before discussing the indices we need to introduce some notations. Assume *K* clusters have been identified in the set of patients, and let Y_{hcj} be the z-score for variable *h* (= 1, 2, …, *p*) on patient *j* (= 1, 2, …, n_c) in cluster *c* (= 1, 2, …, *K*). The total number of patients is $n\ (= n_1 + n_2 + \cdots + n_K)$. The within ($SSW_K$) and between ($SSB_K$) sums of squares of the clusters are given by $SSW_K = \sum_{h=1}^{p}\sum_{c=1}^{K}\sum_{j=1}^{n_c}\left(Y_{hcj} - \bar{Y}_{hc.}\right)^2$ and $SSB_K = \sum_{h=1}^{p}\sum_{c=1}^{K} n_c(\bar{Y}_{hc.} - \bar{Y}_{h..})^2$ with $\bar{Y}_{hc.}$ the average value of the variable *h* in cluster *c*, and $\bar{Y}_{h..}$ the average value of variable *h.* It should be noted that for hierarchical clustering, the clusters are nested, which means that an additional cluster would imply a split of one of the existing clusters $c = 1, 2, \ldots, K$. This means that SSW_1 is equal to the total sums of squares (*SST*) of all the data, with $SST = SSW_K + SSB_K$ for all numbers of clusters $K \geq 1$.

Furthermore, let d_{jk} be the distance between patient *j* and patient *k*, and let $x_{jk}^{c_1c_2}$ be the indicator function defined as follows:

$$x_{jk}^{c_1c_2} = \begin{cases} 1 \text{ if } j \in I_{c_1} \text{ and } k \in I_{c_2} \\ 0 \text{ otherwise,} \end{cases}$$

with I_c the set of patients that are contained in cluster *c*. For convenience, we define $d^{(r)}$ to be the r-th ranked distance among all $n(n-1)/2$ possible distances. Thus $d^{(1)}$ is the smallest distance between any two patients and $d^{\{n(n-1)/2\}}$ is the largest. In hierarchical clustering, clusters can be formed sequentially following the dendrogram from one homogeneous group to *n* clusters at the end. This procedure is followed for all indices and the indices will judge whether the observed set of *K* clusters should be changed to *K*+1 clusters. The way that we walk down the dendrogram is determined by the distance of two clusters (i.e. the length of the arms of the dendrogram), choosing larger distance first.

2.1.1. Beale (B) Index

Beale (1969) proposed an index which allows a significance test for choosing the number of clusters. The existence of an additional cluster is tested by the F-test, and the index is defined by

$$B_K = \frac{(n_c - 2)(SSW_K - SSW_{K+1})}{SSW_{K+1}\{(n_c - 1)2^{2/p} - (n_c - 2)\}}, K \geq 1,$$

where n_c is the number of observations of cluster *c* that was split up to form *K*+1 clusters in total. Note that $SSW_K - SSW_{K+1}$ depends on the data of cluster *c* alone. When B_K is smaller than or equal to the critical value of the F-distribution with *p* and $(n_c - 2)p$ degrees of freedom, cluster *c* is not considered heterogeneous and should not be divided into two clusters (Milligan and Cooper, 1985; Everitt et al., 2011).

2.1.2. Duda and Hart (DH) Index

Duda and Hart (1973) suggested the ratio of the two within sum of squares to decide whether a cluster can be divided into two clusters, i.e.

$$DH_K = \frac{SSW_{K+1}}{SSW_K}, K \geq 1.$$

For large number of observations in cluster *c*, SSW_K is approximately normally distributed with mean $n_c p\sigma^2$ and variance $2n_c p\sigma^4$ and SSW_{K+1} follows approximately a normal distribution with mean $n_c p(1 - 2/\pi p)\sigma^2$ and variance $2n_c p\{1 - 8/(\pi^2 p)\}\sigma^4$. The variance σ^2 represents the variance of the complete population. From these observations the following criterion was proposed to sub-divide whenever the following holds

$$DH_K < \left[1 - \frac{2}{\pi p} - z\sqrt{\frac{2\{1 - 8/(\pi^2 p)\}}{n_c p}}\right],$$

with *z* a standard normal quantile. Milligan and Cooper (1985) suggested to use the value 3.20 for *z* (Everitt et al., 2011; Milligan and Cooper, 1985; Duda and Hart, 1973).

2.1.3. Calinski and Harabasz (CH) Index

Calinski and Harabasz (1974) proposed the CH index for finding the number of clusters:

$$CH_K = \frac{SSB_K/(K-1)}{SSW_K/(n-K)}, K \geq 2.$$

The number of clusters was chosen to be the value of *K* that maximizes CH_K. Note that the CH index always assumes that there are at least two clusters and cannot be used to decide if clustering is needed.

2.1.4. Hartigan (H) Index

Hartigan (1975) originally proposed the *H* index for the number of clusters with K-means clustering. The index is defined by

$$H_K = (n-K-1)\left\{\frac{SSW_K - SSW_{K+1}}{SSW_{K+1}}\right\}, K \geq 2.$$

The term (*n*-*K*-1) is a penalty factor which avoids an increasing monotonicity with an increasing number of clusters. Hartigan (1975) recommended that the number of clusters is the smallest *K* for $H_K \leq 10$. Alternatively, Milligan and Cooper (1985) used $(H_{K+1} - H_K)$ and recommended the number of clusters *K* that maximizes this difference. In this study, we used the criterion of Milligan and Cooper (1985), but this implies that there should always exist at least two clusters even if no subtypes would be present.

2.1.5. C-index

Dalrymple-Alford (1970) proposed the C-index and later Hubert and Levin (1976) revised it. Let d_{jk} be the Euclidean distance, i.e.

$$d_{jk}^2 = \sum_{h=1}^{p} \left(Y_{hj} - Y_{hk}\right)^2,$$

with Y_{hj} the observation on variable *h* and patient *j* ignoring the possible cluster structure in which this patient belong. Within cluster *c*, there are $n_c(n_c-1)/2$ distances to be calculated and the total number of such pairs over all clusters is

$$n_{WK} = \sum_{c=1}^{K} (n_c^2 - n_c)/2.$$

As mentioned before, the total number of pairs in the data set is $n_T = n(n-1)/2$. This total number can be rewritten as $n_T = n_{WK} + n_{BK}$, with $n_{BK} = \sum_{c_1=1}^{K-1} \sum_{c_2=c_1+1}^{K} n_{c_1} n_{c_2}$ the number of pairs that do not belong to the same cluster. If we now define

$S_{WK} = \sum_{c=1}^{K} \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} d_{jk} x_{jk}^{cc}$, $S_{min} = \sum_{r=1}^{n_{WK}} d^{(r)}$, and $S_{max} = \sum_{r=n_{BK}}^{n(n-1)/2} d^{(r)}$, the C-index is defined by

$$C_K = \frac{S_{WK} - S_{min}}{S_{max} - S_{min}}, K \geq 2,$$

under the assumption that $S_{min} \neq S_{max}$. Note that C_K is always an element of the interval [0, 1]. The number of clusters *K* is the number that minimizes C_K. It is always larger than one, which means that C_K cannot be used to decide if clustering should be proceeded.

2.1.6. Cubic Clustering Criteria (CCC)

Sarle (1983) developed a crude test for testing the null hypothesis that data have been sampled from a uniform distribution on a hyperbox against the alternative hypothesis that data have been sampled from a mixture of spherical multivariate normal distribution with equal variances and sampling probabilities. The author compared observed value $R_K^2 = 1 - SSW_K/SST$, the proportion of variance accounted for by the *K* clusters, with an approximation of its expected value $\mathbb{E}R_K^2$ under the assumption that the *K* clusters are generated by a *p*-dimensional uniform distribution. The Cubic Clustering Criteria (CCC) is defined as

$$CCC_K = ln\left[\frac{1 - E(R_K^2)}{1 - R_K^2}\right]\frac{\sqrt{np^*/2}}{[0.001 + E(R_K^2)]^{1.2}}, K \geq 1,$$

and $\mathbb{E}R_K^2$ is defined by

$$\mathbb{E}R_K^2 = 1 - \left[\frac{\sum_{h=1}^{p^*}(n + u_h)^{-1} + \sum_{h=p^*+1}^{p} u_h^2(n + u_h)^{-1}}{\sum_{h=1}^{p} u_h^2}\right][(n - K)^2/n](1 + 4/n),$$

where $u_h = s_h/m$, s_h is the square root of the *h*-th eigen value of *SST*/(*n*-1), $m = (v^*/K)^{1/p^*}$, with $v^* = \prod_{h=1}^{p^*} s_h$ and where p^* is chosen to be the largest integer less than *K* such that u_{p^*} is not less than one. A positive value of CCC_K means that the observed R_K^2 is greater than the expected R_K^2 under the uniform distribution and the cluster structure of the data is different from the uniform partition (i.e. reject the null hypothesis). The number of cluster *K* is determined by the number that maximizes CCC_K.

2.1.7. Krzanowski and Lai (KL) Index

Let $DIFF_K$ denote a scaled difference between the within sum of squares of two sequential clusterings, i.e.

$$DIFF_K = (K - 1)^{2/p} SSW_{K-1} - K^{2/p} SSW_K, K \geq 2.$$

Krzanowski and Lai (1988) argued that under independent uniformly distributed data the sum of squares $K^{2/p} SSW_K$ is constant and independent of the number of clusters *K*. They suggested to calculate the *KL* index by the ratio of two difference measures

$$KL_K = \left|\frac{DIFF_K}{DIFF_{K+1}}\right|, K \geq 2.$$

The proposed number of clusters is the number *K* that maximizes KL_K. The KL_K cannot be used to decide if clustering is needed or not.

2.1.8. Silhouette (S) Index

Kaufman and Rousseeuw (1990) proposed the *S*-index for assessing and estimating the true number of clusters. Let us define the within-cluster mean distance a_j^c as the mean distance of patient *j* to the other patients in cluster *c*, i.e.

$$a_j^c = \frac{1}{n_c - 1}\sum_{K=1} d_{jk} x_{jk}^{cc}.$$

The mean distance of patient *j* in cluster c_1 to the patients in another cluster c_2 is defined by

$$b_j^{c_1c_2} = \frac{1}{n_{c_2}} \sum_{K=1}^{n} d_{jk} x_{jk}^{c_1c_2}.$$

Then the smallest of these mean distances $b_j^{c_1c_2}$ over clusters $c_2 \in \{1, 2, \dots, K\} / \{c_1\}$ is defined by $b_j^c = min\left\{b_j^{c_1c_2} \middle| c_2 \in \{1, 2, \dots, K\} / \{c_1\}\right\}$.

The *silhouette* width of patient *j* in cluster *c* is now given by

$$S_j^c = \frac{b_j^c - a_j^c}{max\{a_j^c, b_j^c\}}.$$

Note that S_j^c is an element of the interval [-1, 1]. A value of S_j^c near one indicates that patient *j* is categorized within the right cluster while a value near minus one indicates that the patient could be better changed to another cluster. The average *silhouette* index for cluster *K* is defined by

$$S_K = \frac{1}{K} \sum_{c=1}^{K} \sum_{j=1}^{n_c} S_j^c / n_c \,, c = 1, 2, \dots, K.$$

The number of clusters is taken as the number that maximizes S_K across the hierarchical formulation of clusters.

2.1.9. Gap/uni and Gap/pc Statistic

Tibshirani, Walther, and Hastie (2001) proposed the Gap statistic as comparing the logarithm of within sums of squares with the expectation of this term under the reference distribution of the data. Therefore, the statistic is defined by

$$GAP_K = \mathbb{E}\{log(SSW_K)\} - log(SSW_K).$$

The expected value $\mathbb{E}log(SSW_K)$ is unknown and it is therefore determined using Monte Carlo simulation from a reference distribution and applying bootstrapping. On the basis of bootstrap sampling with *B* samples $\mathbb{E}log(SSW_K)$ is estimated with $\sum_{b=1}^{B} log(SSW_{K,b}^*) / B$, with $SSW_{K,b}^*$ be the within sum of squares for bootstrap sample *b*.

The number of clusters *K* is determined by the smallest number such that the following holds

$$GAP_K \geq GAP_{K+1} - S_{K+1}, K \geq 1,$$

with $S_K = sd_K\sqrt{(1 + 1/B)}$ is the total standard error and sd_K is given by

$$sd_K^2 = \frac{1}{B} \sum_{b=1}^{B} \left\{ log(SSW_{K,b}^*) - 1/B \sum_{b=1}^{B} log(SSW_{K,b}^*) \right\}^2.$$

There were two choices about the reference distribution for the Gap statistic (Tibshirani et al., 2001). In the first choice, each variable was generated from the uniformly distribution over the range of the observed values for the variable *p*. Determining the number of clusters via this choice was referred to as Gap/uniform or Gap/uni. In the second choice, the variables were sampled from a uniform distribution over a box aligned with the principal components of the centered design matrix. The new design matrix was then back transformed to obtain the reference dataset. This procedure of calculating the number of clusters was referred to as Gap/principal components or Gap/pc.

2.1.10. Weighted Gap (WGap)

Define the sum of pairwise distances between all patients within cluster *c* by

$$\overline{D}_c = \sum_{h=1}^{p}\sum_{j=1}^{n}\sum_{k=1}^{n}\left(Y_{hj} - Y_{hk}\right)^2 x_{jk}^{cc}/[2n_c(n_c-1)],$$

with Y_{hj} the value for patient *j* on variable *h* ignoring the cluster structure. The weighted within sum of squares is defined by $\overline{SSW}_K = \sum_{c=1}^{K}\overline{D}_c$. Yan and Ye (2007) proposed this WGap as an alternative to the original Gap statistic, but followed the exact same approach of Tibshirani *et al.* (2001) to compare it with its expectation. Therefore, the weighted Gap statistic is defined by

$$\overline{GAP}_K = \frac{1}{B}\sum_{b=1}^{B} log\left(\overline{SSW}_{K,b}^{*}\right) - log(\overline{SSW}_K), K \geq 1.$$

Both a WGap/uni and WGap/pc were obtained using the same reference distributions. The number of clusters *K* is determined to be the number that maximizes $\overline{GAP}_K$.

2.1.11. Difference of difference weighted Gap (DD-WGap)

The WGap method was used to test the null hypothesis of one homogeneous group against the alternative of multiple subtypes. If there is more than one clusters, the original Gap statistic and the WGap statistic may have a tendency to overestimate the number of clusters (Dudoit and Fridlyand, 2002; Yan and Ye, 2007). Therefore, the DD-WGap statistic has been proposed by Yan and Ye (2007) to find the best estimate of the number of clusters more efficiently. Let $D\overline{GAP}_K$ denote the difference in two sequential weighted Gap statistics

$$D\overline{GAP}_K = \overline{GAP}_K - \overline{GAP}_{K-1}, K \geq 2.$$

If the data are strongly grouped around *K* (*K* ≥ 2) modes based on $\overline{SSW}_K$, the function is defined by the difference of difference weighted Gap (DD-WGap) function

$$DD\overline{GAP}_K = D\overline{GAP}_K - D\overline{GAP}_{K+1}.$$

The number of clusters *K* is determined by the number that maximizes $DD\overline{GAP}_K$. Both the DD-WGap/uni and the DD-WGap/pc statistics can be computed like with the original Gap statistic.

2.2. Simulation Settings

Simulation designs in existing literature used almost the same approach and their designs were formulated based on hypothetical mean and covariance structures (Tibshirani et al., 2001; Albatineh and Niewiadomska-Bugaj, 2011; Albalate et al., 2011). The present study uses various means and covariance matrices derived from the GROUP study. Five different scenarios were chosen: (i) a single cluster structure, (ii) two clusters with a ratio of cluster sizes being 75% and 25%, (iii) three clusters with a ratio of cluster sizes being 40%, 35% and 25%, (iv) four clusters with a ratio of cluster sizes being 40%, 30%, 20% and 10%, and (v) five clusters with a ratio of cluster sizes set at 35%, 25%, 20%, 15% and 10%. The means and covariance matrices for the eight dimensional cognition variables for the clusters were determined from the GROUP study using K-means clustering with the pre-specified number of clusters described above. Based on these input and the associated ratios of sample sizes we simulated normally distributed data for the five settings described above. For each setting, 1000 datasets of 860 subjects with 8 dimensional cognitive z-scores were generated from the mixture of

multivariate normal distributions. The means and covariance matrices that we used for each setting are presented in Tables A1a to A5b in the Appendix file. We have also simulated the above mentioned five scenarios where we chose only six out of eight variables that contributed to the clusters. The two variables CPT performance index and CPT standard deviation were treated as nuisance variables to see if such setting would alter the performance of the indices. In practice we would not know which of these variables would or would not contribute to subtypes.

Two packages in R, clusterSim and NbClust were used to analyze the GROUP study and simulation studies. RStudio version 0.97.551 (R version 3.0.1) was used throughout the analysis.

3. Results of simulation study

Only seven indices were capable of identifying a single cluster. The other indices assume that clustering should be conducted. For these seven indices, the percentage of simulation runs that a single solution was selected is presented in Table 2.

Table 2: Percentage of identifying a single cluster solution from simulation study

	Simulated number of clusters (%)				
	1	2	3	4	5
B	97.1	91.4	28.3	1.4	0.6
DH	0	0	0	0	0
CCC	0	0	0	0	0
Gap/uni	0	0	5.5	0	0
Gap/pc	98.4	26.3	20.2	8.5	0
WGap/uni	0.6	1.8	0.5	0.1	0
WGap/pc	99.1	53.3	26.8	1.7	1.1

When just one single cluster was simulated, DH, CCC, Gap/uni, and WGap/uni were incapable of identifying just one cluster. They always seem to indicate incorrectly a multiple clusters solution. Contrary, B, Gap/pc, and WGap/pc correctly identified a single cluster solution with 97.1%, 98.4% and 99.1% respectively. However, when two and three clusters were simulated, B frequently incorrectly predicted a single cluster solution: 91.4% for 2 clusters and 28.3% for 3 clusters. Gap/pc seems to do better, because it incorrectly chooses 26.3% and 20.2% single cluster solutions when two and three clusters were simulated respectively. The Gap/pc seems to be a good index to answer the question if clustering should be conducted. The WGap/pc on the other hand, performs somewhere in between the performance of B and Gap/pc. Combining Gap/pc and WGap/pc improves the performance. If they both indicate that a single cluster solution is present, they correctly identify a single cluster with 97.5%. When two or three clusters are present they incorrectly predict a single cluster with 20.5% and 6.4%, respectively. For larger number of clusters these percentages vanish.

Table 3 shows the percentages of correctly identifying the exact number of simulated clusters from 1000 simulated datasets for each of the methods (columns (a)). Additionally, the percentage of datasets for which the identified number of clusters deviates no more than one cluster from the simulated number of clusters is also specified (columns (b)). This percentage informs us how frequent a method would predict within a range of K-1, K, K+1 clusters, when we intentionally

simulated *K* clusters. We believe that this is a measure of closeness or stability, which could be relevant too when the frequencies of predicting the correct number of clusters is relatively low. If the correct number is missed, how far away are the predictions from the correct number. For instance, compare CCC with KL on three simulated clusters. The CCC predicts three clusters with 12%, while KL predicts three clusters with 42.1%. Clearly KL seems to perform better than CCC for three clusters. This is only partly true, since KL predicts 2, 3 or 4 in 60.8% and predicts 1, 5, 6 or more clusters still with almost 40%. CCC never predicts 1, 5, 6 or more clusters and provides therefore a stable estimate on the number of clusters. So choosing between KL and CCC is not as simple.

The results demonstrate that the DH index predicts the simulated number of clusters within plus or minus one (*K*-1, *K*, *K*+1) quite well (82.3% - 95.4%) over the range of 2 to 5 clusters. When it comes to identifying the exact number, DH predicts this with about 50% when 3 to 5 clusters would be present.

Table 3: Percentage of identifying the exact number of simulated clusters and within a range of just one cluster from simulation study

Indices	Simulated number of clusters (%)							
	2		3		4		5	
	(a)*	(b)**	(a)	(b)	(a)	(b)	(a)	(b)
B	8.6	100.0	1.0	71.7	0	0.1	0	0
DH	28.1	82.3	43.4	95.4	50.0	89.9	52.6	89.7
CH	100.0	100.0	0.2	100.0	0	0.3	0	0
H	97.1	100.0	4.6	100.0	73.2	99.7	10.5	33.2
C	41.6	55.6	13.7	37.1	18.9	49.5	26.0	49.6
CCC	100.0	100.0	12.0	100.0	0.1	0.7	0.4	0.7
KL	40.3	54.2	42.1	60.8	46.4	61.2	14.3	34.9
S	100.0	100.0	0	100.0	0	0	0	0
Gap/uni	1.8	47.6	20.4	54.2	55.7	84.0	35.3	85.8
Gap/pc	55.5	99.8	30.7	56.1	70.0	87.3	18.4	45.8
WGap/uni	42.0	83.6	35.5	90.7	23.0	67.2	7.9	36.3
WGap/pc	45.0	99.9	13.5	73.0	4.7	26.0	1.3	8.9
DD-WGap/uni	82.7	91.3	5.3	94.4	0.3	1.8	0.2	1.3
DD-WGap/pc	36.4	46.9	13.2	71.0	2.1	7.4	0.8	3.7

*Predicting exact number of simulated clusters, **Predicting the exact number of simulated clusters within a range of just one cluster: {*K*-1, *K*, *K*+1}.

The H index performs better (99.0% to 100%) than DH in predicting the number of clusters within plus or minus one when the simulated number of clusters is less than 5, but it fails dramatically when 5 clusters are simulated (33.2%). Predicting the exact number of clusters with H is low when 3 and 5 underlying clusters are present (4.6% and 10.5%), while this percentage is quite good for 2 and 4 clusters (97.1% and 73.2%). On the other hand, Gap/pc predicts the number of clusters within the range of plus minus one better than DH at 2 clusters. However, DH and H predict the number of clusters better than Gap/pc when 3 and 4 underlying clusters are present. At 5 clusters, Gap/pc is better than H, but worse than DH. Some other indices, such as CH, CCC, and S have perfect performances when 2 clusters are present, but do not perform very well at other settings. They seem to have a preference to predict always 2 clusters, whatever the underlying cluster structure. Furthermore, the WGap/pc recovers quite well at 2 and 3 clusters, but failed at 4 and 5 clusters. It does not do much better than its originator Gap/pc index. B is not very good in predicting

the exact number of clusters for more than one cluster. Both C and KL indices are quite robust over the whole range of the number of clusters, but they have only a medium level performance. Interestingly enough, the Gap/uni seems to perform better when higher number of clusters is involved, although it never performs very well. Finally, the WGap/uni, DD-WGap/uni, and DD-WGap/pc, which represent the latest developments in indices, did not do better than earlier developed indices in our simulations. In addition, when we treated two out of eight variables as nuisance variables for all five simulation designs, the results did not change dramatically and our conclusions remained the same (data not shown). This may indicate that the performance of the indices is not that much affected by nuisance variables. When we would first decide if clustering is needed based on the Gap/pc and WGap/pc together, the results are not changed much either (data not shown).

4. Conclusions

Clustering is an explorative analysis and one of the major challenges is to determine the number of clusters in a complex heterogeneous dataset. Although complex statistical methods are available (using mixture models), relative simple and straightforward methods like K-means clustering are used most frequently. They are often supported with hierarchical clustering to help identifying the number of clusters and select good initial centroids for K-means clustering. Therefore, we investigated the most promising indices for detecting the correct number of clusters on the basis of hierarchical clustering (with Ward's agglomerative method). The indices were investigated on (i) how well they would decide between a single and multiple cluster solution and (ii) on how well they can predict the number of clusters in a multi cluster solution. We simulated clusters based on a real case study of patients with schizophrenia and their neurocognitive measured variables. This complements on the performances of the indices from literature, since the indices were mainly studied with artificial low dimensional simulated data. Although our results support earlier studies, we also found opposite results.

Milligan and Cooper (1985) demonstrated that the B index performed relatively well in identifying the number of clusters when three or more clusters were simulated and less for two clusters. They noted that the B index is an appropriate index when the clusters are well separated and spherical (Beale, 1969; Tibshirani et al., 2001). Although our clusters were also spherical, they were mostly not well separated and this explained the opposite result we found with the B index. In our study, the B index demonstrated a good performance for a single cluster and two clusters, but it failed to detect the correct number of clusters at four and five clusters. Milligan and Cooper (1985) already mentioned the good performance of DH. They also demonstrated that DH was particularly capable of identifying high numbers of clusters and they ranked DH as the second best index, among 30 indices. We demonstrated that DH was most consistent among all the indices we studied with high performances of recovering the number of clusters within a range of one. The CH index was considered the best in the simulation study of Milligan and Cooper (1985) and performed good in almost all investigations of Tibshirani et al. (2001). We confirmed that CH performed well up to 3 clusters, but it failed completely at higher number of clusters. This index seems to have a preference for choosing 2 clusters (in all settings). Tibshirani et al. (2001) showed that the H index did not

perform well at 3 and 4 clusters when the dimension of variables was low (2 or 3 variables). This was supported by the simulations of Albatineh and Niewiadomska-Bugaj (2011). Milligan and Cooper (1985) ranked the H index at place 16 among 30 indices. However, for our eight-dimensional cognition outcomes, the H index did particularly well at 2 to 4 clusters. The C index was originally developed for binary outcome data, but Milligan and Cooper (1985) demonstrated that it performed adequately for continuous outcome too. They showed that it performed well, except for 2 cluster solutions. They ranked this index as the third best index among 30 indices. In our simulation the index was consistent for all cluster solution, but the recovery of clusters was only moderate. It predicted the number of clusters within a range of one with approximately 50%. The CCC seems to have a preference for choosing 2 clusters in our simulation, similar as the CH index. It failed to predict the number of clusters for higher number of cluster solutions and underestimates the number of clusters. Milligan and Cooper (1985) and Boone (2011) demonstrated an opposite observation, namely that the CCC index would over-estimate the number of clusters. This could possibly be explained by the fact that our clusters are not clearly separated. We found that the KL index was consistent across different numbers of clusters, but it had only a medium performance. This supported the results of Albatineh and Niewiadomska-Bugaj (2011) and Marriott (1971), who showed that approximately 40 to 50% of the clusters were adequately identified with the KL index. It was however less than the results of Tibshirani et al. (2001), who demonstrated higher performances for 2 to 4 clusters. The S index performed quite well in the study of Albatineh and Niewiadomska-Bugaj (2011) and for 3 clusters in the study of Tibshirani et al. (2001). We demonstrated that the S index selected frequently two clusters under almost all settings, similar to the CH index. This implies that the S index is less suitable for higher number of clusters and explains its good performance for lower number of clusters. Our result on the Gap/pc index is in line with the results of the originator Tibshirani et al. (2001). The Gap/pc performs best at single clusters and lower number of clusters and less at higher number of clusters. This distinction in performance though was not observed by Albatineh and Niewiadomska-Bugaj (2011). The Gap/uni seemed to perform in our simulation reasonable well at 4 and 5 clusters, but not very well at lower number of clusters. This was contrary to the results of Tibshirani et al. (2001) and Albatineh and Niewiadomska-Bugaj (2011), who showed unsatisfactory results at higher number of clusters. Yan and Ye (2007) introduced several alternative indices that were based on the ideas of Gap, but their results could not be reproduced. We did not demonstrate that any of the WGap/uni, WGap/pc, DD-WGap/uni, and DD-WGap/pc performed systematically better than Gap/uni or Gap/pc. Only at three clusters did these indices performed better, but in other settings it was less than Gap/uni or Gap/pc.

Note that we also simulated settings where only six out of eight variables contributed to the subtypes, but these settings did not change our conclusions on the performance of the indices. They performed almost identical to the setting where all eight variables determined the subtypes. However, we focused only on one particular distance measure for identifying the number of clusters. It may be possible that some of the indices may provide different results if other distances or dissimilarity measures are used. Furthermore, our study investigated the proposed indices using sequential stopping criteria for the hierarchical formation of clusters, i.e. the contribution of a new cluster was evaluated with respect to the previous number of clusters, wherever the next cut in the

tree would happen. If the next set of clusters did not contribute, the number of clusters was determined. This means that the formation of clusters in the dendrogram cuts the branches horizontally. Alternatively, cutting branches can also be performed dynamically, which means that at certain trees in the dendrogram the splitting of clusters may continue, while at other parts of the dendrogram trees are not split up further. It would be of interest to find out whether dynamic cutting would improve the performance of certain indices. Furthermore, the criterion for comparing two sequential sets of hierarchical clusters is not changed with the number of previous comparisons (i.e. multiple testing issues). For instance, the z-value in the criterion for the DH index was selected at 3.20, which corresponds with a two-sided significance level of 0.0014. Instead, the criterion could be altered every time a next comparison is conducted, starting with a lower value for a single cluster and slowly rising to higher significance levels when the number of comparisons increases. A more strict criterion at the single cluster could improve the performance of the DH index for single clusters, but also for other indices. Finally, the indices may also be compared with alternative methods for identifying the number of clusters, such as the Bayesian Information Criterion that is applied to clustering methods that uses maximum likelihood. We studied only the indices, since they fit better with the relative simple but frequently used methods of clustering.

In conclusion, we found that the DH, H, and Gap/pc were the best performing indices in our simulation study based on eight-dimensional outcome variables taken from a real case study of schizophrenic patients. They predicted the simulated number of clusters within the range of one cluster with high probabilities. The DH index was most consistent, while Gap/pc in combination with WGap/pc is capable of answering the question if a multiple cluster solution is present.

References

Albalate, A., Suendermann, D., Minker, W. (2011). On cluster validation for detecting the number of clusters in a data set. *International Journal on Artificial Intelligence Tools* 20:941-953.

Albatineh, A.N., Niewiadomska-Bugaj, M. (2011). MCS: A Method for Finding the Number of Clusters. *Journal of classification* 28:184-209.

Beale, E.M.L. (1969). *Cluster analysis*. UK: London: Scientific Control Systems.

Blyler, C.R., Gold, J.M., Iannone, V.N., Buchanan, R.W. (2000). Short form of the WAIS-III for use with patients with schizophrenia. *Schizophrenia research* 46:209-215.

Boone, D.S. (2011). Determination of the number of clusters in a data set: A stopping rule X clustering algorithm comparison. *International Journal of Strategic Decision Sciences* 2(4):1-13.

Borgen, F.H., Barnett, D.C. (1987). Applying cluster-analysis in counseling psychology research. *Journal of counseling psychology* 34:456-468.

Brand, N., Jolles, J. (1985). Learning and retrieval rate of words presented auditorily and visually. *The Journal of general psychology* 112:201-210.

Bray, J.H., Maxwell, S.E., Cole, D. (1995). Multivariate statistics for family psychology research. *Journal of family psychology* 9:144-160.

Calinski, R.B., Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics* 3:1-27.

Chen, W.J., Faraone, S.V. (2000). Sustained attention deficits as markers of genetic susceptibility to schizophrenia. *American Journal of Medical Genetics* 97:52-57.

Clatworthy, J., Buick, D., Hankins, M., Weinman, J., Horne, R. (2005). The use and reporting of cluster analysis in health psychology: a review. *British journal of health psychology* 10:329-358.

Cornblatt, B.A., Keilp, J.G. (1994). Impaired attention, genetics, and the pathophysiology of schizophrenia. *Schizophrenia bulletin* 20:31-46.

Dalrymple-Alford, E. (1970). Measurement of clustering in free recall. *Psychological bulletin* 74:32-34.

Dawes, S.E., Jeste, D.V., Palmer, B.W. (2011). Cognitive profiles in persons with chronic schizophrenia. *Journal of clinical and experimental neuropsychology* 33:929-936.

Duda, R.O., Hart, P.E. (1973). *Pattern classification and scene analysis*. New York, USA: New York: Wiley.

Dudoit, S., Fridlyand, J. (2002). A prediction-based resampling method for estimating the number of clusters in a dataset. *Genome biology* 3:RESEARCH0036.

Everitt, B.S., Landau, S., Leese, M., Stahl, D. (2011). *Cluster Analysis.* UK: Wiley Series in Probability and Statistics, A John Wiley and Sons, Ltd.

Fernández, A., Gómez, S. (2008). Solving Non-uniqueness in Agglomerative Hierarchical Clustering Using Multidendrograms. *Journal of Classification* 25:43-65.

Gordon, A.D. (1999). *Classification*. London, UK: Chapman & Hall/CRC Monographs on Statistics & Applied Probability.

Hartigan, J.A. (1975). *Clustering Algorithms. A Wiley Publication in Applied Statistics, John Wiley & Sons, New York.*

Hartigan, J.A., Wong, M.A. (1979). A K-Means Clustering Algorithm. *Applied Statistics* 28:100-108.

Henry, D.B., Tolan, P.H., Gorman-Smith, D. (2005). Cluster analysis in family psychology research. *Journal of family psychology : JFP : journal of the Division of Family Psychology of the American Psychological Association (Division 43)* 19:121-132.

Hubert, L.J., Levin, J.R. (1976). A general statistical framework for assessing categorical clustering in free recall. *Psychological Bulletin* 83:1072-1080.

Jablensky, A. (2006). Subtyping schizophrenia: implications for genetic research. *Molecular psychiatry* 11:815-836.

Joyce, E.M., Roiser, J.P. (2007). Cognitive heterogeneity in schizophrenia. *Current opinion in psychiatry* 20:268-272.

Kaufman, L., Rousseeuw, P.J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. USA: New York: Wiley-Interscience.

Korver, N., Quee, P.J., Boos, H.B.M., Simons, C.J.P., de Haan, L., GPOUPinvestigators, Investigators, G. (2012). Genetic Risk and Outcome of Psychosis (GROUP), a multi-site longitudinal cohort study focused on gene-environment interaction: objectives, sample characteristics, recruitment and assessment methods. *International Journal of Methods in Psychiatric Research* 21:205-221.

Krzanowski, W.J., Lai, Y.T. (1988). A criterion for determining the number of groups in a data set using sum-of-squares clustering. *Biometrics* 44:23-34.

Lin, S.H., Liu, C.M., Liu, Y.L., Shen-Jang Fann, C., Hsiao, P.C., Wu, J.Y., Hung, S.I., Chen, C.H., Wu, H.M., Jou, Y.S., Liu, S.K., Hwang, T.J., Hsieh, M.H., Chang, C.C., Yang, W.C., Lin, J.J., Chou, F.H., Faraone, S.V., Tsuang, M.T., Hwu, H.G., Chen, W.J. (2009). Clustering by neurocognition for fine mapping of the schizophrenia susceptibility loci on chromosome 6p. *Genes, brain and behavior* 8:785-794.

Mandara, J. (2003). The typological approach in child and family psychology: a review of theory, methods, and research. *Clinical child and family psychology review* 6:129-146.

Marriott, F.H. (1971). Practical problems in a method of cluster analysis. *Biometrics* 27:501-14.

Meijer, J., Simons, C.J., Quee, P.J., Verweij, K., GROUP Investigators. (2012). Cognitive alterations in patients with non-affective psychotic disorder and their unaffected siblings and parents. *Acta Psychiatrica Scandinavica* 125:66-76.

Milligan, G.W., Cooper, M.C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika* 50:159-179.

Quee, P.J., Alizadeh, B.Z., Aleman, A., van den Heuvel, E.R., GROUP Investigators. (2014). Cognitive subtypes in non-affected siblings of schizophrenia patients: characteristics and profile congruency with affected family members. *Psychological medicine* 44:395-405.

Quee, P.J., van der Meer, L., Bruggeman, R., de Haan, L., Krabbendam, L., Cahn, W., Mulder, N.C., Wiersma, D., Aleman, A. (2011). Insight in psychosis: relationship with neurocognition, social cognition and clinical symptoms depends on phase of illness. *Schizophrenia bulletin* 37:29-37.

Sarle, W.S. (1983). *Cubic clustering criterion (Tech. Rep. A-108)*. USA: Cary, N.C.: SAS Institute.

Silver, H., Shmoish, M. (2008). Analysis of cognitive performance in schizophrenia patients and healthy individuals with unsupervised clustering models. *Psychiatry research* 159:167-179.

Stinissen, J., Willems, P.J., Coetsier, O., Hulsman, W.L.L. (1970). *Manual of the Dutch Edition of the WAIS.* . The Netherlands: Lisse, The Netherlands: Swets & Zeitlinger.

Tibshirani, R., Walther, G., Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society.Series B, Statistical methodology* 63:411-423.

Yan, M., Ye, K. (2007). Determining the number of clusters using the weighted gap statistic. *Biometrics* 63:1031-7.

Appendix

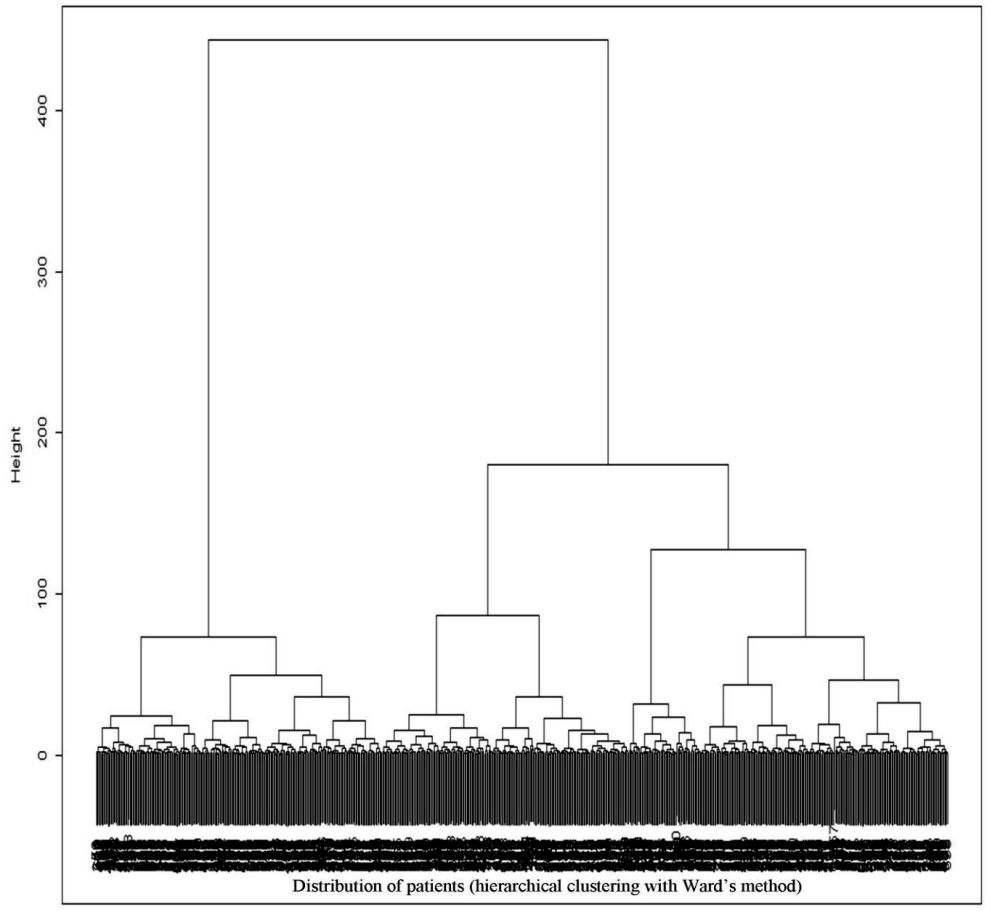


Figure A1: Dendrogram for schizophrenia patients using cognitive variables from GROUP

Table A1a: Means of cognition variables for the single cluster solution

Cluster	CPTpi	CPTsd	Block	Symbol	Calc	Info	WLT_I	WLT_D
1	-0.58	0.73	-0.49	-1.10	-0.80	-0.51	-0.92	-0.68

Table A1b. Variance-covariance matrix of cognition variables for the single cluster solution

Variables	CPTpi	CPTsd	Block	Symbol	Calc	Info	WLT_I	WLT_D
CPTpi	1.33	-0.78	0.27	0.42	0.28	0.20	0.31	0.23
CPTsd		1.62	-0.34	-0.48	-0.40	-0.27	-0.34	-0.26
Block			1.33	0.52	0.66	0.69	0.41	0.38
Symbol				1.05	0.54	0.50	0.41	0.34
Calc					1.19	0.76	0.48	0.40
Info						1.24	0.48	0.42
WLT_I							1.20	0.85
WLT_D								1.03

CPTpi: Continuous performance test index, CPTsd: Standard deviation of CPT, Block: Block design, Symbol: Digit symbol coding, Calc: Arithmetic/calculation, Info: Information, WLT_I: Word learning task immediate recall, and WLT_D: Word learning task delayed recall.

Table A2a: Means of cognition for the two clusters solution

Clusters	CPTpi	CPTsd	Block	Symbol	Calc	Info	WLT_I	WLT_D
1	-1.05	1.35	-1.25	-1.73	-1.6	-1.3	-1.59	-1.24
2	-0.19	0.22	0.12	-0.59	-0.15	0.14	-0.38	-0.23

Table A2b: Variance-covariance matrices of cognition variables for the two clusters solution

Cluster	Variables	CPTpi	CPTsd	Block	Symbol	Calc	Info	WLT_I	WLT_D
1	CPTpi	1.35	-0.63	0.08	0.24	0.06	-0.08	0.07	0.04
1	CPTsd		1.63	0.01	-0.21	-0.05	0.10	-0.04	0.01
1	Block			0.92	0.14	0.19	0.20	0.03	0.08
1	Symbol				0.64	0.17	0.10	0.09	0.08
1	Calc					0.66	0.24	0.07	0.06
1	Info						0.80	0.07	0.08
1	WLT_I							0.83	0.50
1	WLT_D								0.68
2	CPTpi	0.99	-0.48	-0.10	0.12	-0.11	-0.12	0.05	0.00
2	CPTsd		1.04	0.08	-0.13	0.06	0.15	0.02	0.03
2	Block			0.82	0.12	0.15	0.21	-0.03	0.01
2	Symbol				0.80	0.10	0.09	0.05	0.04
2	Calc					0.69	0.25	0.02	0.02
2	Info						0.68	0.04	0.04
2	WLT_I							0.83	0.59
2	WLT_D								0.86

CPTpi: Continuous performance test index, CPTsd: Standard deviation of CPT, Block: Block design, Symbol: Digit symbol coding, Calc: Arithmetic/calculation, Info: Information, WLT_I: Word learning task immediate recall, and WLT_D: Word learning task delayed recall.

Table A3a: Means of cognition variables for the three clusters solution

Clusters	CPTpi	CPTsd	Block	Symbol	Calc	Info	WLT_I	WLT_D
1	-0.25	0.30	0.35	-0.40	0.15	0.40	-0.15	0.00
2	-0.30	0.40	-0.65	-1.20	-1.0	-0.75	-1.15	-0.90
3	-1.60	0.20	-1.60	-2.10	-1.90	-1.55	-1.80	-1.40

Table A3b: Variance-covariance matrices of cognition variables for the three clusters solution

Clusters	Variables	CPTpi	CPTsd	Block	Symbol	Calc	Info	WLT_I	WLT_D
1	CPTpi	1.00	-0.55	0.00	0.18	0.00	0.00	0.09	0.00
1	CPTsd		1.21	0.00	-0.25	0.00	0.07	-0.05	-0.05
1	Block			0.56	0.07	0.05	0.07	-0.03	0.00
1	Symbol				0.81	0.06	0.00	0.00	-0.04
1	Calc					0.49	0.09	-0.06	-0.09
1	Info						0.42	0.00	0.00
1	WLT_I							0.72	0.47
1	WLT_D								0.72
2	CPTpi	0.81	-0.41	-0.14	0.09	-0.11	-0.16	0.00	0.00
2	CPTsd		0.81	0.18	-0.03	0.00	0.12	0.00	0.04
2	Block			1.00	0.00	0.04	0.14	-0.21	-0.16
2	Symbol				0.49	0.00	-0.03	-0.06	-0.06
2	Calc					0.64	0.14	-0.07	-0.06
2	Info						0.81	-0.11	-0.11
2	WLT_I							0.73	0.41
2	WLT_D								0.64
3	CPTpi	1.44	-0.23	-0.10	0.09	-0.05	-0.26	-0.06	0.00
3	CPTsd		1.56	0.21	0.00	0.09	0.32	0.12	0.11
3	Block			0.72	0.03	0.16	0.18	0.00	0.11
3	Symbol				0.56	0.14	0.16	0.00	0.10
3	Calc					0.56	0.26	0.07	0.10
3	Info						0.72	0.08	0.11
3	WLT_I							0.90	0.57
3	WLT_D								0.72

CPTpi: Continuous performance test index, CPTsd: Standard deviation of CPT, Block: Block design, Symbol: Digit symbol coding, Calc: Arithmetic/calculation, Info: Information, WLT_I: Word learning task immediate recall, and WLT_D: Word learning task delayed recall.

Table A4a: Means of cognition variables for the four clusters solution

Clusters	CPTpi	CPTsd	Block	Symbol	Calc	Info	WLT_I	WLT_D
1	-1.60	1.90	-1.70	-2.15	-2.00	-1.70	-1.95	-1.55
2	-1.05	1.42	-0.02	-1.15	-0.45	-0.10	-1.10	-0.80
3	0.00	-0.05	0.45	-0.30	0.20	0.45	0.05	0.15
4	0.10	-0.03	-1.05	-1.20	-1.30	-1.05	-1.00	-0.85

Table A4b: Variance-covariance matrices of cognition variables for the four clusters solution

Clusters	Variables	CPTpi	CPTsd	Block	Symbol	Calc	Info	WLT_I	WLT_D
1	CPTpi	1.50	-0.25	-0.10	0.10	0.10	-0.25	-0.10	0.00
1	CPTsd		1.65	0.05	-0.10	-0.10	0.20	0.15	0.10
1	Block			0.60	0.00	0.05	0.10	-0.10	0.05
1	Symbol				0.50	0.10	0.05	-0.05	0.10
1	Calc					0.35	0.15	0.05	0.00
1	Info						0.70	0.00	0.05
1	WLT_I							0.90	0.55
1	WLT_D								0.70
2	CPTpi	0.60	-0.10	0.10	0.05	-0.10	0.00	-0.10	-0.15
2	CPTsd		0.95	-0.10	-0.10	-0.10	-0.15	0.20	0.25
2	Block			0.70	0.05	0.00	0.00	-0.10	-0.15
2	Symbol				0.65	0.10	0.10	-0.10	-0.10
2	Calc					0.70	0.15	-0.05	-0.05
2	Info						0.70	-0.10	-0.05
2	WLT_I							0.60	0.35
2	WLT_D								0.65
3	CPTpi	0.80	-0.30	-0.05	0.10	-0.05	0.00	-0.10	-0.10
3	CPTsd		0.70	0.05	-0.10	0.05	0.10	0.20	0.15
3	Block			0.55	0.05	0.05	0.10	-0.05	0.05
3	Symbol				0.90	0.10	0.00	-0.10	-0.10
3	Calc					0.55	0.10	-0.10	-0.10
3	Info						0.50	0.05	0.05
3	WLT_I							0.65	0.50
3	WLT_D								0.80
4	CPTpi	0.85	-0.30	0.10	0.15	0.10	0.00	-0.15	-0.05
4	CPTsd		0.60	-0.15	-0.10	-0.15	-0.10	0.10	0.10
4	Block			1.00	0.15	-0.05	0.00	-0.15	-0.15
4	Symbol				0.50	0.05	0.05	-0.05	-0.05
4	Calc					0.65	0.20	0.00	-0.05
4	Info						0.70	-0.05	-0.05
4	WLT_I							0.85	0.50
4	WLT_D								0.70

CPTpi: Continuous performance test index, CPTsd: Standard deviation of CPT, Block: Block design, Symbol: Digit symbol coding, Calc: Arithmetic/calculation, Info: Information, WLT_I: Word learning task immediate recall, and WLT_D: Word learning task delayed recall.

Table A5a. Means of cognition variables for the five clusters solution

Clusters	CPTpi	CPTsd	Block	Symbol	Calc	Info	WLT_I	WLT_D
1	-1.85	2.05	-1.70	-2.25	-2.05	-1.65	-2.10	-1.70
2	-1.10	1.80	-0.20	-1.30	-0.65	-0.25	-1.00	-0.65
3	-0.25	0.20	0.40	-0.30	0.10	0.45	0.45	0.60
4	0.10	-0.25	0.30	-0.60	-0.10	0.10	-1.05	-1.00
5	-0.15	0.20	-1.35	-1.35	-1.55	-1.35	-1.15	-0.90

Table A5b: Variance-covariance matrices of cognition variables for the five clusters solution

Clusters	Variables	CPTpi	CPTsd	Block	Symbol	Calc	Info	WLT_I	WLT_D
1	CPTpi	1.50	-0.15	-0.10	0.10	0.05	-0.25	-0.15	-0.10
1	CPTsd		1.70	-0.05	-0.10	-0.05	0.15	0.25	0.10
1	Block			0.65	0.05	0.05	0.10	-0.10	-0.05
1	Symbol				0.50	0.10	0.10	-0.05	0.05
1	Calc					0.35	0.15	0.00	-0.05
1	Info						0.65	0.00	-0.05
1	WLT_I							0.85	0.50
1	WLT_D								0.60
2	CPTpi	0.55	-0.10	0.10	0.10	-0.15	-0.05	0.00	-0.15
2	CPTsd		0.85	0.00	0.10	0.05	-0.10	0.10	0.10
2	Block			0.70	0.05	0.05	0.00	-0.10	-0.10
2	Symbol				0.65	0.10	0.05	0.05	0.10
2	Calc					0.80	0.15	0.05	0.00
2	Info						0.70	0.00	0.00
2	WLT_I							0.55	0.30
2	WLT_D								0.60
3	CPTpi	0.90	-0.25	-0.15	0.15	-0.05	0.10	0.10	0.00
3	CPTsd		0.70	-0.05	-0.15	0.10	0.10	0.10	0.00
3	Block			0.65	0.10	0.15	0.15	-0.10	0.05
3	Symbol				0.90	0.10	0.15	-0.10	-0.10
3	Calc					0.55	0.15	-0.05	-0.05
3	Info						0.55	0.05	-0.10
3	WLT_I							0.40	0.20
3	WLT_D								0.45
4	CPTpi	0.95	-0.35	-0.05	0.05	-0.15	-0.15	0.15	0.25
4	CPTsd		0.50	0.10	0.00	0.05	0.15	-0.15	-0.20
4	Block			0.65	0.10	-0.05	-0.05	-0.05	-0.10
4	Symbol				0.75	0.05	0.05	-0.10	-0.10
4	Calc					0.60	0.20	-0.05	-0.10
4	Info						0.65	-0.05	0.10
4	WLT_I							0.50	0.20
4	WLT_D								0.50
5	CPTpi	0.75	-0.25	-0.05	0.10	0.05	-0.05	-0.05	0.05
5	CPTsd		0.60	-0.10	-0.10	-0.10	-0.10	-0.10	0.00
5	Block			0.70	0.15	-0.10	0.00	-0.15	0.00
5	Symbol				0.55	0.15	-0.05	0.15	0.10
5	Calc					0.60	0.10	0.00	-0.10
5	Info						0.65	0.00	0.05
5	WLT_I							0.90	0.55
5	WLT_D								0.70

CPTpi: Continuous performance test index, CPTsd: Standard deviation of CPT, Block: Block design, Symbol: Digit symbol coding, Calc: Arithmetic/calculation, Info: Information, WLT_I: Word learning task immediate recall, and WLT_D: Word learning task delayed recall.

CHAPTER 3

Long-term cognitive trajectories in patients with schizophrenia and their unaffected siblings

Md. Atiqul Islam
Frank D. van Es
Tesfa D. Habtewold
Piotr J. Quee
Edwin R. van den Heuvel
Behrooz Z. Alizadeh[‡]
Richard Bruggeman[‡]
GROUP Investigators

[‡]Equal contribution

Submitted to Psychological Medicine

Abstract

Background: Heterogeneity of psychosis is reflected in the cognitive functioning of patients with psychosis and their unaffected siblings. This study assessed the heterogeneity and stability of cognition in patients with schizophrenia and their unaffected siblings, by forming separate longitudinal trajectories. Next, we aimed to predict cognitive subtypes of siblings by their probands.

Methods: Assessments were conducted at baseline, three and six years in 1,119 patients, 1,059 siblings and 586 controls from the Genetic Risk and Outcome of Psychosis (GROUP). Group-based trajectory modeling was applied to identify trajectories, and clustered multinomial logistic regression analysis was used for prediction modeling. Cognition was based on a composite measure of eight neurocognitive tests.

Results: Five cognitive trajectories were identified for patients, ranging from impaired to high performance. Four trajectories were found for siblings, ranging from moderately impaired to high performance. These distinct subtypes were stable and persisting over time. Siblings had higher risks to perform moderately impaired if patients perform mildly altered, moderately and severely impaired, compared to the combination of normal and high performance, with odds ratios 2.21 (95% CI 1.05-4.64), 5.70 (2.77–11.70) and 10.07 (4.15–24.44) respectively. The familial correlation coefficient between pairs of index patients and their siblings was 0.27 (P=0.003).

Conclusion: Cognitive performance of patients and of their unaffected siblings is heterogeneous and stable over time. Cognitive subtypes of patients significantly predicted subtypes of siblings. The profiling approach used in the current study is suitable for usage in future genetic studies, as well as in predicting functional and clinical outcomes.

Keywords: cognition; cognitive trajectory; heterogeneity, psychosis; siblings; subtypes

1. Introduction

Schizophrenia spectrum disorders consist of multiple symptom dimensions, caused by the interaction of genetic, environmental, and internal factors (van Os et al., 2010). One of these dimensions is cognitive impairment and it has been demonstrated that it is a predictor of symptomatic and functional outcome (*e.g.* working activity, daily living activity) (Heinrichs and Zakzanis, 1998; Green et al., 2000; Faber et al., 2011; Nuechterlein et al., 2014; Harvey, 2014).

Siblings of psychotic patients have been found to be a heterogeneous group with respect to cognition as well (Keri and Janka, 2004; Meijer et al., 2012). They exhibit subtle cognitive deficits in different domains, such as sustained attention, working memory, verbal memory and verbal fluency (Kremen et al., 1994; Chen and Faraone, 2000; Trandafir et al., 2006; Faraone et al., 1999; Krukow et al., 2017). From a clinical perspective, these subtle cognitive changes are thought of as markers for endophenotypic vulnerability to schizophrenia (Gur et al., 2007). The expression of psychotic vulnerability measured by neurocognition is higher in family members of patients, as compared to the normal population (Vollema and Postma, 2002). First-degree biological relatives of patients are approximately 10-fold risk of developing schizophrenia compared to the normal population (Kendler and Diehl, 1993). This indicates that cognitive impairment in schizophrenia cannot be solely attributed to the influence of disease-related factors, such as psychotic episodes, hospitalization, and unemployment or medication effects. Thus, studying relatives offers a unique possibility to unravel pathogenetic mechanisms, excluding confounding factors such as pharmacotherapy or life-style.

In First Episode Psychosis (FEP), patients' cognitive impairment seems stable during multiple years after onset (Barder et al., 2013; Townsend and Norman, 2004; Albus et al., 2006; Rodriguez-Sanchez et al., 2008; Leeson et al., 2011; Bozikas and Andreou, 2011). It has been suggested that their cognitive performance is stable over time regarding attention, verbal memory, and executive functioning (Faraone et al., 1995; Faraone et al., 1999). However, these studies of cognitive performance over time are based on mean values that do not take into account heterogeneity. In a previous study we demonstrated that subtype classification of unaffected siblings of patients with schizophrenia supports the evidence of heterogeneity in cognitive function (Quee et al., 2014).

In this study, we aimed to unravel the heterogeneity of neurocognition in patients and their siblings by classifying their neurocognitive performance time profile. Additionally, we aimed to predict cognitive subtypes of siblings by subtypes of patients.

2. Methods

2.1. Study design, setting and participants

The current study was performed within the framework of the Genetic Risk and Outcome of Psychosis (GROUP) project, a longitudinal multi-center cohort study in the Netherlands and Belgium. A group of outpatients and inpatients with psychotic disorder between 16 and 50 years were recruited. Siblings were asked to participate if they had at least one participating sibling with a non-affective psychotic disorder according to DSM-IV (American Psychiatric Association, 2000). Siblings were included if they (i) were between 16 and 50 years, (ii) had a good command of Dutch language and (iii) had no lifetime psychotic disorder. For controls, the inclusion and exclusion criteria were the same as for siblings. The procedure of recruitment, informed consent, approval by the accredited

Medical Ethics Review Committee (METC) and population characteristics have been described in detail elsewhere (Korver et al., 2012). Between April 2004 and December 2013, participants were assessed at study entry, three and six years thereafter.

The full GROUP sample consisted of 1,119 patients, 1,059 unaffected siblings and 586 healthy controls at baseline. For standardization of the neurocognitive tests by age and gender, 586 subjects with seemingly no mental or somatic disease were included as a control group. We included all samples based on eight cognitive measures (see the details in section 2.3) from study entry, three and six years thereafter.

2.2. Sample size calculation

There was no formal sample size calculation for our intended analysis of trajectory modeling. However, Formann (1984) estimated the sample size for latent class analysis. According to Formann, the minimal sample size to include for latent class analysis was no less than 2^k cases with k the number of variables (Formann, 1984). In our case, we would need at least $2^k = 2^8 = 256$ subjects in our analysis, which we satisfy broadly.

2.3. Assessment of Neurocognition

Task selection was based on cognitive domains that have been shown to be impaired in schizophrenia (Nuechterlein et al., 2004). The cognitive battery has been described in details elsewhere (Meijer et al., 2012). This study focused on the neurocognitive measures which were related to outcome in an earlier study (Meijer et al., 2012). Supplementary Table S1 shows the list of cognitive domains, their corresponding tests, and outcome measures.

We calculated composite neurocognitive scores based on the following eight neurocognitive measures: the Continuous Performance Test (CPT), Word Learning Task immediate recall and delayed recall, WAIS-III Symbol Substitution, Information, Arithmetic, and Block Design (Supplementary Table S1). For the CPT, a measure was calculated for the performance index called 'CPT performance', and the reaction time variability called 'CPT variance' (See Supplementary Table S1). Subsequently, linear regression analyses (i.e. controls specific age and gender on each cognitive measure) were conducted for each time point. The scores of control subjects were used to obtain age and gender adjusted z-scores for both patients and siblings on all eight neurocognitive tests. Finally, composite scores for the overall cognitive functioning of patients and siblings respectively were computed by averaging z-scores of all eight tests.

2.4. Assessments of socio-demographics and clinical variables

Educational degree was evaluated as a continuous variable according to Verhage (Verhage, 1964). Level of premorbid functioning was assessed using Premorbid Adjustment Scale (PAS) (Cannon-Spoor et al., 1982). Positive and negative symptoms of schizotypy were measured using the Structured Inventory for Schizotypy-Revised (SIS-R) (Vollema and Ormel, 2000). Frequency and distress of psychotic experiences were measured by the Community Assessment of Psychic Experiences (CAPE) (Brenner et al., 2007). The symptom severity of patients was assessed with the 30-items five-factor Positive And Negative Syndrome Scale (PANSS) (Kay et al., 1987; Lancon et al., 2000). Age of onset of

psychosis was also measured for patients. Other socio-demographics variables were age of participants, gender and ethnicity.

2.5. Data analysis and Statistical Modeling

2.5.1. Descriptive Statistics

The socio-demographics characteristics at baseline for the controls, siblings, and patients were compared using univariate analyses. For gender and ethnicity, Pearson's Chi-square test was used to test the difference between groups (controls, siblings, and patients). Due to the family structure data, linear mixed effects models were applied on all continuous variables (e.g. socio-demographics, clinical and individual cognitive tests) to test for differences between the groups (controls, siblings, and patients). Family was taken as a random effect variable. The method of Maximum Likelihood (ML) was used to estimate the model parameters and Type-III (overall) tests of fixed effects were used to test for differences between groups. Significant differences were followed by pair-wise post-hoc comparisons between groups. Additionally, Pearson's correlation coefficients between cognitive performances of both patients and siblings were computed to explore the possible predictive relationship.

2.5.2. Trajectory modeling

Group-based trajectory modeling (Nagin, 1999; Jones et al., 2001; Nagin and Odgers, 2010; Niyonkuru et al., 2013) was conducted in order to identify clusters of patients and siblings, separately with similar patterns of composite neurocognitive scores over time. Longitudinal measurements of composite scores were treated as dependent variables and follow-up time (baseline, three and six years) as independent variables. The first order linear and second order quadratic polynomial model was fitted assuming that individual differences in trajectories could be summarized by a finite set of polynomial functions of time. To determine the number of clusters, a sequential approach was applied where the number of clusters is increased by one. The less complex model (i.e. less trajectory groups) was compared with the complex model (i.e. more groups) using the Bayesian Information Criterion (BIC) (Schwarz, 1978) and logged Bayes factor (2*ΔBIC), where ΔBIC = BIC (complex) – BIC (less complex) (Wit et al., 2012; Kass and Raftery, 1995) and where the logged Bayes factor (2*ΔBIC) would indicate trivial (0-2), positive (2-6), strong (6-10) or very strong (>10) evidence for the null hypothesis that the less complex model is the best fit. At first, a single quadratic polynomial trajectory model was examined. If the quadratic term was not significant, the model was re-run with linear trajectory to determine the BIC value. If the quadratic component in one trajectory model was significant, the quadratic two-trajectory model was performed. Next, the BIC value of the appropriate two-trajectory model compared to the BIC value of the appropriate one-trajectory model. The process was repeated with an increasing number of trajectories until the best fit model was found, as determined by comparing the BIC values and logged Bayes factor (Andruff et al., 2009). The trajectory modeling will be stopped at the moment when the ΔBIC becomes negative value. The ML estimation can handle missing data in the cognition score when it would satisfy the missing at random mechanism (Little and Rubin, 2002). We also checked the dropout model which includes a logistic model of dropout probability of cognitive functioning per period to see the dependency of

the cognitive functioning of patients and their siblings with the function of three time points. For each group, 0 = constant rate, 1 = depends on the previous response (baseline), 2 = depends on the two previous responses (baseline and three years).

2.5.3. Comparison among cognitive subtypes on all baseline characteristics

Differences in all continuous variables between the subtypes of patients and siblings respectively were investigated using linear mixed effect models taking into account familial relationship. Type-III (overall) tests of fixed effects were used to test for differences between cognitive subtypes. If these were significant, pair-wise comparisons between subtypes were investigated with Dunnet's method (taking the most normal cognitive profile as reference group). For gender and ethnicity, Pearson's Chi-square test was used to test the difference between trajectory groups of patients and siblings respectively.

2.5.4. Cognitive subtypes of Patients predict subtypes of siblings

Our hypothesis of the sibling-patient analysis was that subtype of patient predicted the subtype of sibling. We considered sibling subtype (multi-category) as dependent and patient subtype (multi-category) as independent variables. Concordant, discordant and Somers' D statistic (Somers, 1962) were computed on the pairs of subtypes of patients and siblings. A pair of subtypes of patients, subtypes of siblings-pairs was said to be concordant if the larger value of subtypes of patients was paired with the larger value of subtypes of siblings, and was said to be discordant if the larger value of subtypes of patients was paired with the smaller value of subtypes of siblings. Somers' D of subtypes of siblings with respect to subtypes of patients was defined as the difference between the two conditional probabilities of concordance and discordance. In the modeling, given the family structure of the data (as siblings-patients belong within the same family) clustered multinomial logistic regression was conducted taking into account family as a random effect. PROC NLMIXED in SAS (Statistical Analysis System) was applied to determine the predictive relationship between subtypes of patients and siblings. An adaptive Gaussian quadrature with 10 quadrature points was specified to integrate out the random effect of the likelihood function (Kuss and McLerran, 2007; de Rooij and Worku, 2012) and to estimate the parameters (i.e. subtypes of patients) and their standard errors. The intra-cluster correlation coefficient (ICC) was calculated to estimate the familial correlation between pairs of unaffected siblings and schizophrenic index patients in the same family. The ICC was calculated as

$ICC_{Family} = var(family)/(var(family) + \pi^2/3)$, where, *var(family)* is the variance of random effect and π is 3.14159.

A two-tailed test at P<0.05 is considered as statistical significant throughout the analyses. All analyses were performed using Statistical Analysis System (SAS), version 9.4.

3. Results

3.1. Descriptive of study population

Differences between patients, siblings, and controls were significant (P<0.001) on socio-demographic, cognitive and clinical variables at baseline (Table 1). Pair-wise comparisons revealed that differences were significant between patients and controls as well as between siblings and controls. On all measures, patients and siblings displayed poorer performances than controls. Additionally, performances on cognitive outcomes of patients with schizophrenia were found to be significantly lower than the performances of their unaffected siblings (Table 1).

Pearson's correlation coefficients between cognitive performances were significant for all cognitive tests in patients and in siblings; except for Block Design, Arithmetic and Information with CPT performance of siblings (Supplementary Table S2-S3).

Table 1: Comparison of baseline characteristics for controls, siblings, and patients*.

Variable/Group	Group			Overall group difference	Pair-wise group comparison
	1. Controls (n=586)	2. Siblings (n=1,059)	3. Patients (n=1,119)		
Age	30.42 (10.58)	27.84 (8.28)	27.58 (7.94)	F=22.8, P<0.001	2<1, 3<1
Gender, % male	45.90	45.51	76.14	χ^2=254.1, P<0.001	2<1<3
Education (Verhage)[a]	5.41 (1.78)	5.07 (2.11)	4.04 (2.05)	F=128.5, P<0.001	3<2<1
Ethnicity, % Dutch	92.12	83.24	79.22	χ^2=45.2, P<0.001	3<2<1
IQ, Estimated[b]	109.75 (15.08)	102.76 (15.60)	94.99 (16.12)	F=185.6, P<0.001	3<2<1
PAS, overall score[c]	1.13 (0.59)	1.13 (0.66)	1.98 (0.88)	F=439.0, P<0.001	2<3, 3<1
SIS-R[d]					
Positive	0.31 (0.35)	0.38 (0.42)	…	F=15.4, P<0.001	1<2
Negative	0.24 (0.22)	0.27 (0.26)	…	F=9.0, P=0.002	1<2
Age of onset	…	…	23.69 (7.59)	…	…
PANSS 5-factor					
Positive	…	…	13.90 (6.55)	…	…
Negative	…	…	15.00 (6.64)	…	…
Disorganization	…	…	16.77 (6.27)	…	…
Excitement	…	…	12.05 (4.05)	…	…
Emotional distress	…	…	15.82 (5.73)	…	…
CAPE (Positive dimension)[e]					
PE Frequency	0.19 (0.17)	0.21 (0.20)	0.67 (0.49)	F=566.5, P<0.001	2<3, 3<1
PE Distress	0.43 (0.45)	0.46 (0.48)	1.26 (0.69)	F=531.0, P<0.001	2<3, 3<1
Cognitive Performance					
CPT performance[f]	246.36 (54.78)	243.70 (57.86)	220.82 (62.14)	F=52.0, P<0.001	3<2, 1<3
CPT variance (ms)[g]	72.76 (28.28)	75.80 (28.44)	92.99 (36.51)	F=108.4, P<0.001	2<3, 3<1
Block Design[h]	46.55 (14.16)	44.87 (15.07)	40.42 (16.99)	F=39.2, P<0.001	3<2<1
Digit Symbol[i]	84.01 (14.58)	79.21 (15.39)	65.41(16.27)	F=379.9, P<0.001	3<2<1
Arithmetic[j]	15.32 (4.16)	13.86 (4.43)	12.27 (4.79)	F=93.1, P<0.001	3<2<1
Information[k]	18.84 (4.67)	16.83 (5.22)	16.77 (5.47)	F=31.5, P<0.001	2<1, 3<1
Immediate Recall[l]	28.47 (5.37)	26.89 (5.77)	22.94 (6.09)	F=224.5, P<0.001	3<2<1
Delayed Recall[m]	9.74 (2.70)	9.33 (2.64)	7.53 (2.87)	F=181.6, P<0.001	3<2<1

*Table presents means (SD) or %; Empty space (…) means no measurments in the respective group (row/column wise); [a]Education (Verhage): range 0 (no education), 3-5 (school diploma) to 8 (university degree); [b]IQ: Wechsler Adult Intelligence Scale-III (WAIS-III), short form; [c]PAS: Premorbid Adjustment Scale; [d]SIS-R: Structured Inventory for Schizotypy – Revised; [e]CAPE: Community Assessment for Psychic Experiences; PE frequency and distress: Frequency of positive psychotic experiences and amount of distress of these PE; [f]CPT performance: Continuous Performance Test HQ, performance index, [g]CPT variance (ms): CPT-HQ variance in reaction time (ms); [h]Block Design: WAIS-III Block Design; [i]Digit Symbol: WAIS-III Digit Symbol Substitution Test; [j]Arithmetic: WAIS-III Arithmetic; [k]Information: WAIS-III Information; [l]Immediate recall: Word Learning Task (WLT) immediate recall; [m]Delayed recall: WLT Delayed recall. For the PAS, higher scores reflect poorer premorbid adjustment. [1]Control, [2]Sibling and [3]Patient; Pair-wise group comparison explains which group better or worse in terms of measurements.

3.2. Trajectory modeling and longitudinal course

To identify cognitive trajectories of patients and their unaffected siblings over time, group-based trajectory modeling was applied. A five-group cognitive trajectory model for patients and a four-group trajectory model for siblings were very strongly favoured for overall cognitive scores (z-scores) according to the smallest BIC and the logged Bayes factor (Supplementary Table S4). In patients, the value of logged Bayes factor 17.34 (>10) favouring the five-cluster model over a six-cluster model. In siblings, the value of logged Bayes factor 94.62 (>10) considering the four-trajectory model over a five-cluster model (Supplementary Table S4). Parameter estimates of linear and quadratic polynomial time functions of trajectory modeling including dropout models are presented in the Supplementary Table S5 and S6.

The figure 1a and 1b displayed the changes of cognitive trajectories of patients and their siblings on z-scores of composite cognitive measure over six years. The cognitive trajectories in patients were labeled as ‘severely impaired’, ‘moderately impaired’, ‘mildly altered’, ‘normal’ and ‘high performer’. Similarly, the trajectories in siblings were labeled as ‘moderately impaired’, ‘mildly altered’, ‘normal’ and ‘high performer’. All initial profiles of composite neuro-cognition for patients and siblings were stable over time (Figure 1a-1b). Severe and moderate groups were identified based on a broad-based cognitive impairment of, on average, about 1 SD below the normal across a range of composite cognition score. A large group of patients (26.7%) and the largest group of sibling (37.6%) displayed normal cognitive functioning. The majority of patients (30.4%) and a large group of siblings (25.1%) showed mild cognitive alterations across all time points. Then, a group of patients (28.4%) and a smaller group of siblings (13.0%) manifested moderate cognitive impairment. The latter group of siblings showed steady and slight improvement of performance over time. Severe impairment was only seen in a small group of patients (10.7%).

The smallest patient group (3.8%) and a substantial group of siblings (24.2%) performed higher cognitive functioning, compared to the mean level of the normal cognitive subtype. Patients in this cluster showed a slight decline over time (Figure 1a), whereas the pattern was stable for siblings (Figure 1b).

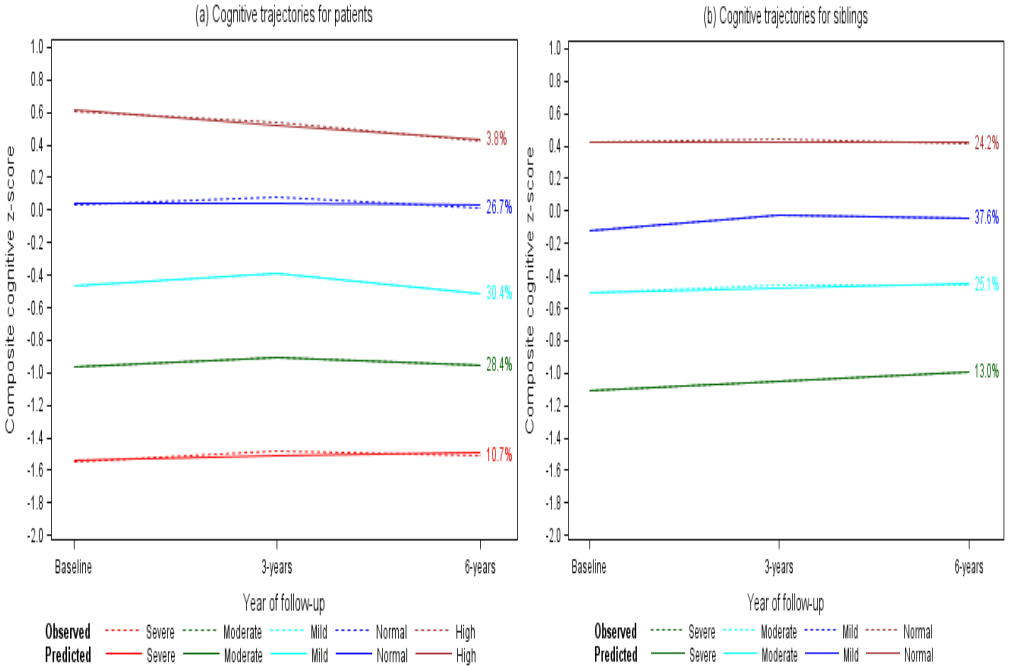


Figure 1: (a) Cognitive trajectories for patients (n=1,119), (b) Cognitive trajectories for siblings (n=1,059)

3.3. Comparison of cognitive subtypes on all baseline characteristics

Differences between cognitive subtypes of siblings were significant with regard to age, gender, education, ethnicity, IQ, PAS overall score and SIS-R positive symptoms, CAPE frequency of psychotic experiences (PE) and all cognitive performances except CPT variance (Table 2). Siblings with a moderate impaired profile were lower educated and poorer cognitive performance than those with a normal profile. Similarly, mild altered group of siblings had lower education, lower IQ and poorer cognitive performance compared to the performance of normal profile (Table 2).

Cognitive subtypes of patients were compared on all cognitive measures and on demographic and clinical variables (Table 3). Pair-wise comparisons of subtypes revealed that patients with a severely impaired profile performed poorer on all cognitive measures. Besides, significant differences were found between profiles of severe impairment and normal performance, on measures of education, age of onset of psychosis, IQ, the PAS overall score, and PANSS negative, disorganization and excitement score. Patients with a moderate impaired profile were lower educated, lower IQ, poorer premorbid functioning, poorer symptoms of PANSS on all five factors, and poorer cognitive performance than those with a normal profile. Difference between the mildly altered and normal group of patients were found on measures of education, IQ and all cognitive measures except CPT variance.

Table 2: Characteristics, main effects and pair-wise comparisons for the four subtypes of non-affected siblings (n=1,059)*

Variable/Trajectory	Cognitive trajectory group				Overall trajectory group difference	Pair-wise comparison**
	1. High (n=254)	2. Normal (n=413)	3. Mild (n=260)	4. Moderate (n=132)		
Age	28.6 (7.7)	27.9 (8.0)	27.8 (8.9)	26.3 (8.9)	F=5.6, P=0.001	2<1
Gender, % male	55.5	43.6	41.5	40.2	χ^2=14.1, P=0.003	2<1; 2>4
Education (Verhage)[a]	6.1 (1.8)	5.3 (2.0)	4.5 (2.0)	3.5 (2.1)	F=62.2, P<0.001	2<1,3,4
Ethnicity, % Dutch	87.4	85.7	79.5	74.8	χ^2=14.1, P=0.003	2<1; 2>3,4
IQ, Estimated[b]	120.7 (9.9)	104.9 (9.5)	92.1 (7.9)	81.9 (6.7)	F=684.7, P<0.001	2<1,3,4
PAS, overall score[c]	0.9 (0.6)	1.1 (0.7)	1.2 (0.6)	1.3 (0.7)	F=13.6, P<0.001	2<4, 2<1
SIS-R[d]						
Positive	0.3 (0.4)	0.4 (0.4)	0.4 (0.4)	0.5 (0.5)	F=4.2, P=0.006	2<4
Negative	0.3 (0.3)	0.3 (0.2)	0.3 (0.3)	0.3 (0.3)	F=2.5, P=0.108	
CAPE (Positive dimension)[e]						
PE Frequency	0.2 (0.2)	0.2 (0.2)	0.2 (0.2)	0.3 (0.3)	F=4.3, P=0.006	2<4
PE Distress	0.4 (0.4)	0.5 (0.5)	0.5 (0.5)	0.6 (0.5)	F=2.4, P=0.074	
Cognitive Performance						
CPT performance[f]	254.9 (58.9)	245.3 (49.8)	242.4 (59.7)	220.9 (68.3)	F=9.3, P<0.001	4<2
CPT variance (ms)[g]	70.9 (26.0)	74.6 (26.7)	80.0 (30.3)	80.2 (32.3)	F=4.8, P=0.003	
Block Design[h]	56.3 (8.4)	48.2 (12.4)	37.1 (13.8)	28.1 (11.9)	F=208.3, P<0.001	2<1,3,4
Digit Symbol[i]	89.7 (11.9)	81.4 (13.0)	72.7 (14.4)	65.5 (13.8)	F=121.7, P<0.001	2<1,3,4
Arithmetic[j]	17.7 (2.6)	14.9 (3.2)	11.5 (3.6)	8.2 (3.1)	F=318.5, P<0.001	2<1,3,4
Information[k]	21.8 (3.5)	17.5 (4.0)	14.1 (3.8)	10.6 (3.6)	F=299.9, P<0.001	2<1,3,4
Immediate Recall[l]	31.4 (4.4)	27.3 (4.9)	25.0 (5.0)	20.7 (4.8)	F=159.0, P<0.001	2<1,3,4
Delayed Recall[m]	11.3 (2.1)	9.5 (2.4)	8.5 (2.3)	6.6 (1.8)	F=113.2, P<0.001	2<1,3,4

*Table presents mean (standard deviation) or %; **Pair-wise trajectory comparison: always compared with normal group using Dunnett's adjustment; (1, 3, 4) ordering is lower to higher difference (with respect to mean or proportion) than normal. [a]Education (Verhage): range 0 (no education), 3-5 (school diploma) to 8 (university degree); [b]IQ: Wechsler Adult Intelligence Scale-III (WAIS-III), short form; [c]PAS: Premorbid Adjustment Scale; [d]SIS-R: Structured Inventory for Schizotypy – Revised (higher scores in PAS and SIS-R reflect poorer outcomes); [e]CAPE: Community Assessment for Psychic Experiences; PE frequency and distress: Frequency of positive psychotic experiences and amount of distress of these PE; [f]CPT performance: Continuous Performance Test HQ, performance index, [g]CPT variance (ms): CPT-HQ variance in reaction time (ms); [h]Block Design: WAIS-III Block Design; [i]Digit Symbol: WAIS-III Digit Symbol Substitution Test; [j]Arithmetic: WAIS-III Arithmetic; [k]Information: WAIS-III Information; [l]Immediate recall: Word Learning Task (WLT) immediate recall; [m]Delayed recall: WLT Delayed recall; Cognitive trajectory: high (1) to worst (4).

Table 3: Characteristics, main effects and pair-wise comparisons for the five subtypes of patients (n=1,119)*

Variable/Trajectory	Cognitive trajectory group					Overall trajectory group difference	Pair-wise comparison**
	1. High (n=31)	2. Normal (n=290)	3. Mild (n=377)	4. Moderate (n=312)	5.Severe (n=109)		
Age	31.4 (8.3)	27.3 (7.2)	28.0 (8.0)	27.3 (8.4)	26.8 (8.0)	F=2.5, P=0.051	2<1
Gender, % male	74.2	77.9	74.3	77.2	75.2	χ^2=1.6, P=0.816	
Education (Verhage)[a]	5.8 (1.9)	4.9 (1.8)	4.3 (2.0)	3.3 (1.9)	2.6 (1.8)	F=49.0, P<0.001	2<1,3,4,5
Ethnicity, % Dutch	90.3	87.2	79.9	75.1	64.5	χ^2=30.6, P<0.001	2<1; 2>3,4,5
Age of Onset	26.2 (8.1)	23.0 (6.5)	23.2 (7.9)	22.9 (7.9)	23.0 (7.3)	F=1.4, P=0.253	
IQ, Estimated[b]	127.8 (10.7)	110.5 (10.4)	95.7 (8.6)	83.3 (7.7)	72.9 (6.7)	F=615.6, P<0.001	2<1,3,4,5
PAS, overall score[c]	1.8 (1.0)	1.8 (0.8)	1.9 (0.8)	2.1 (1.0)	2.3 (0.8)	F=9.7, P<0.001	2<5; 2<4
PANSS 5-factor[d]							
Positive	14.0 (8.2)	13.4 (5.9)	13.1 (5.9)	15.0 (7.1)	14.8 (7.7)	F=4.0, P=0.003	2<4
Negative	15.1 (6.4)	13.6 (5.9)	14.3 (6.4)	16.1 (7.0)	18.0 (7.1)	F=11.1, P<0.001	2<5; 2<4
Disorganization	16.0 (6.8)	14.8 (4.9)	15.7 (5.6)	18.3 (6.6)	21.3 (7.1)	F=30.5, P<0.001	2<5; 2<4
Excitement	12.8 (5.2)	11.5 (3.5)	11.5 (3.7)	12.7 (4.4)	13.2 (4.5)	F=7.0, P=0.001	2<5; 2<4
Emotional distress	16.0 (5.7)	15.4 (5.4)	15.2 (5.5)	16.7 (5.8)	16.4 (6.8)	F=3.3, P=0.024	2<4
Cognitive Performance							
CPT performance[e]	256.1 (45.7)	240.9 (52.2)	225.8 (59.6)	210.7 (56.5)	170.0 (78.8)	F=30.1, P<0.001	2<3,4,5
CPT variance (ms)[f]	84.2 (35.3)	83.1 (31.9)	90.1 (36.0)	101.8 (37.2)	106.3 (38.9)	F=13.4, P<0.001	2<4,5
Block Design[g]	60.4 (8.8)	53.5 (10.9)	42.2 (14.3)	30.7 (13.8)	20.6 (11.0)	F=197.6, P<0.001	2<1,3,4,5
Digit Symbol[h]	85.2 (14.6)	76.2 (14.6)	66.5 (12.5)	57.8 (12.9)	47.5 (11.7)	F=138.7, P<0.001	2<1,3,4,5
Arithmetic[i]	18.8 (2.5)	16.2 (3.2)	13.0 (3.8)	9.1 (3.3)	6.7 (2.7)	F=264.22, P<0.001	2<1,3,4,5
Information[j]	24.6 (2.3)	21.1 (3.4)	17.4 (4.3)	13.6 (4.3)	9.9 (3.5)	F=243.9, P<0.001	2<1,3,4,5
Immediate Recall[k]	32.4 (3.8)	27.1 (4.7)	23.4 (4.7)	19.9 (4.8)	15.9 (5.0)	F=173.2, P<0.001	2<1,3,4,5
Delayed Recall[l]	12.1 (1.8)	9.4 (2.4)	7.6 (2.4)	6.2 (2.2)	4.6 (2.0)	F=142.7, P<0.001	2<1,3,4,5

*Table presents mean (standard deviation) or %; **Pair-wise trajectory comparison: always compared with normal group using Dunnett's adjustment; (1, 3, 4, 5) ordering is lower to higher difference (with respect to mean or proportion) than normal. [a]Education (Verhage): range 0 (no education), 3-5 (school diploma) to 8 (university degree); [b]IQ: Wechsler Adult Intelligence Scale-III (WAIS-III), short form; [c]PAS: Premorbid Adjustment Scale; [d]PANSS: Positive and Negative Syndrome Scale (higher scores in PAS and PANSS reflect poorer outcomes); [e]CPT performance: Continuous Performance Test HQ, performance index, [f]CPT variance (ms): CPT-HQ variance in reaction time (ms); [g]Block Design: WAIS-III Block Design; [h]Digit Symbol: WAIS-III Digit Symbol Substitution Test; [i]Arithmetic: WAIS-III Arithmetic; [j]Information: WAIS-III Information; [k]Immediate recall: Word Learning Task (WLT) immediate recall; [l]Delayed recall: WLT Delayed recall; Cognitive trajectory: high (1) to worst (5).

3.4. Cognitive subtypes of patients in relation to subtypes of siblings

In sibling-patient pair analysis, we generated 1,070 pairs of affected and unaffected siblings. The number of pairs was more than 1,059 because we paired multiple unaffected siblings with their single affected sibling or multiple affected siblings with their single unaffected sibling within a family. The contigency table of the subtypes of patients and siblings is presented in the Supplementary Table S7. Somers' *D* determined the association between cognitive subtypes of patients and subtypes of siblings amongst 1070 sib-pairs. A positive value of Somers' *D* (0.29) indicates that the siblings have better cognitive scores than their probands. Thus, being ill means a deterioration of cognitive performances or a lower cognitive performance means a higher risk of becoming ill.

Since the cell frequency of a moderate group of siblings and high cognitive performance group of patients was empty (Supplementary Table S7), i.e. moderate group of siblings did not have probands who were in high performer groups, we combined high cognitive performer group of patients to the normal cognitive group. The combined normal and high performer group of patients was considered as the reference group.

Table 4 presents the predictive relationship between cognitive sybtypes of patients and siblings. Overall, cognitive subtypes of patients significantly predicted the sibling subtype. Here, we present a risk of an unaffected sibling to be grouped in any of mildly altered, moderately impaired and severely impaired groups given the cognitive trajectory of their corresponding affected sibling. The familial correlation i.e. the intra-class correlation coefficient between pairs of unaffected siblings and index patients in the same family accounted 27% (P=0.003) of total variation.

We observed that a sibling in the group of severely impaired patients was at risk (odds ratio (OR) 2.56, 95% CI 1.26–5.18) of having mild alterations of unaffected siblings. This estimate was 1.83 (95% CI 1.12 – 2.98) for affected siblings of moderately impaired patients. However, mild cognitive alterations of patients did not predict mild alterations of his/her unaffected sibling (OR 0.86, P=0.545; Table 4a).

Siblings showed to be at risk to perform moderately impaired if their probandsperformed mildly altered, moderately or severely impaired, compared to the combined normal and high performer group of affected siblings, with OR's of 2.21 (95% CI 1.05-4.64), 5.70 (95% CI 2.77–11.70) and 10.07 (95% CI 4.15–24.44) respectively (Table 4b).

Subsequently, unaffected siblings had low risks to have high cognitive performance if patients were severely (OR 0.26, 95% CI 0.09-0.63) or moderately (OR 0.39, 95% CI 0.24-0.64) impaired; and to have mildly altered (OR 0.37, 95% CI 0.24-0.59) cognitive functioning, when the probands performed in the combined normal and high performance group (Table 4c). In general, of patients who performed severely impaired, unaffected siblings were likely to show moderate to mild alterations of cognitive functioning.

Table 4: Parameter estimates of subtypes of patients on the subtype of sibling (n=1,070 pairs)*.

Cognitive performance of patient subtype	Prediction of cognitive subtypes in siblings	
	OR (95% C.I)	P-value
a. Siblings with mild alterations		
Intercept	0.56 (0.38 - 0.81)	0.002
Severe impairment of affected sib	2.56 (1.26 – 5.18)	0.009
Moderate impairment of affected sib	1.83 (1.12 – 2.98)	0.015
Mild alterations of affected sib	0.86 (0.53 – 1.41)	0.545
b. Siblings with moderate impairment		
Intercept	0.11 (0.06 - 0.21)	<0.001
Severe impairment of affected sib	10.07 (4.15 – 24.44)	<0.001
Moderate impairment of affected sib	5.70 (2.77 – 11.70)	<0.001
Mild alterations of affected sib	2.21 (1.05 - 4.64)	0.036
c. Siblings with high performance		
Intercept	1.24 (0.90 - 1.71)	0.186
Severe impairment of affected sib	0.24 (0.09 - 0.63)	0.003
Moderate impairment of affected sib	0.39 (0.24 - 0.64)	<0.001
Mild alterations of affected sib	0.37 (0.24 - 0.59)	<0.001
Random effect variance and ICC	Estimate (95% C.I)	P-value
Variance (Family)	1.19 (0.12 – 2.26)	0.029
ICC	0.27 (0.09 – 0.44)	0.003

*Reference category for patients is the combination of normal and high cognitive performance, and for siblings normal performance. OR = Odds Ratio; C.I = Confidence Interval; ICC = Intra-class correlation coefficient.

4. Discussion

The aim of this study was two-fold. Firstly, we examined ways to unravel the heterogeneity of neurocognition, by classifying patients and siblings respectively into different subgroups based on the course of composite overall cognition scores over time, separately from each other, based on eight neurocognitive tests. After determining the meaningful subgroups of patients and siblings, we aimed to predict the cognitive subtypes of siblings by subtypes of patients within a family using sibling-patient analysis. Most of the longitudinal studies published in literature investigated the cognitive trajectories over time in patients only. Therefore, to the best of our knowledge, this is the first large-scale study which included cognitive trajectory modeling for both patients and their unaffected siblings.

At baseline, cognitive performance in first-degree relatives of patients with schizophrenia was worse, compared to controls. In line with previous studies (Braff et al., 2007; Krabbendam et al., 2001; Quee et al., 2014), we demonstrated that the performance of siblings were between those of patients and controls. This indicates a parallel between cognitive performance and familial liability.

Trajectory modeling demonstrated a five-trajectory model for patients and a four-trajectory model for siblings that were stable over time. Approximately 70% of patients demonstrated poor cognitive performance (Severe + moderate + mild) (Figure 1a), which was similar to other schizophrenia studies (Thompson et al., 2012; Irani et al., 2011; Szoke et al., 2008). Two of these studies (Barder et al., 2013; Barder et al., 2013), used continuous scale of cognitive measures, demonstrated that cognitive domains remained stable over five to ten years after first episode psychosis. However, they did not classify sub-groups of cognitive performance. In our study, a sub-group of 10.7% patients showed severe cognitive impairment and it was stable over a six-year period.

On the other hand, in the sibling model, a total of 38% (Moderate + mild) exhibited lower cognitive performance than the normal performing subtype over the full period of six years. Both moderate and mild alterations groups of siblings were stable over time (Figure 1b). This result is similar to an earlier analysis of a cross-sectional study done by our research group (Quee et al., 2014). In this study, mild to severely impaired groups of patients- and siblings performed at least -1SD below normal cognition and their trajectories were stable over time.

Studying cognitive trajectories in siblings may provide insight into individuals who are at risk for psychosis. Siblings in the moderate impaired group were more often an ethnic minority, younger, of low IQ, and with a higher level of psychotic experiences (Table 1). This study rated 13% of siblings and 10.7% of patients as moderate to severely impaired. As was found during earlier analysis (Quee et al., 2014) and other studies (Palmer et al., 1997; Kremen et al., 2000), clinical outcomes (e.g. PANSS) and premorbid functioning was poorer in these subgroups (Table 2). Identifying meaningful trajectories lends support to the clinical notion that cognitive deficits are moderate to severely impaired across several domains of cognitive battery, and these impairments are the reasons of disabilities in functioning (Bowie and Harvey, 2006; Keefe and Harvey, 2012).

The positive value for the measure of Somers' *D* (0.29) from the sib-pair analysis showed that the proband is more likely to show low cognitive performance than the sibling, indicating that the disease has a direct impact on cognitive performance. On the other hand, the cognitive subtype of the patient significantly predicted the cognitive subtype of the sibling within the family. The poorer the cognitive profile of the patient, the better it predicted (OR 10.07) the profile of a more cognitively impaired sibling. Moreover, siblings in high performer group are less predicted when their probands were severely impaired (OR 0.24).

We found a familial correlation of 27%, which is high compared to other complex diseases such as depression or bipolar disorder. Of note, this correlation did not take into account the genetic and other environmental factors as it takes only the familial effect. Literature found that neurocognitive impairment is a known inherited form of familial schizophrenia (Brzustowicz et al., 2000; Brzustowicz et al., 2004; Husted et al., 2009). Some family-based studies have already found high rates of cognitive impairment in unaffected relatives of individuals with schizophrenia than in the general population (Snitz et al., 2006). Husted et al. (2009) investigated the heritability of seven distinct neurocognitive measures for schizophrenia, and found significant heritability between 31 to 62% (Husted et al., 2009). Therefore, we would have expected stronger familial effect if we should have taken genetic and other environmental factors into account, since cognitive impairment is not the only factor related to the symptoms of schizophrenia, it is one of the proxies of schizophrenia.

Neurocognitive measures have been proposed as reliable endophenotypic markers of liability for schizophrenia, as cognitive deficits are transmitted within families of patient with schizophrenia (Gur et al., 2007; Krabbendam et al., 2001). The presented subtypes of neurocognition can also be regarded as candidate endophenotypic markers, as they are associated with socio-demographic and clinical variables. Additionally, cognitive subtypes are stable over time and are an inherited form of psychotic disorder. Several studies have shown that cognitive function as a composite measure is a better predictor of functional outcome than any single cognitive test (Bilder et al., 2000; Mohamed et

al., 1999). Further studies are needed to evaluate the additive value of cognitive subtypes of siblings and patients in predicting functional, clinical outcomes and quality of life.

The strength of the study is that we included a large number of patients with schizophrenia, their unaffected siblings, and healthy controls. We included multiple patients and siblings within the same family and their measurements over time yielded an effective study which jointly investigated on the cognitive trajectory of patients and siblings. Moreover, we provided a hint of prediction of sibling's cognitive impairment by the different cognitive profiles of schizophrenia patients. In the methodological perspective, we used group-based trajectory modeling (Nagin, 2014) to identify the significant long-term cognitive trajectory considering drop-out modeling within the same model. Other studies did not take into account the drop-out modeling of the dependency of cognitive functioning with the function of several time points separately when they identified groups (Barder et al., 2013; Barder et al., 2013; Thompson et al., 2012). Some limitations should also be mentioned in this study. There was selection bias in data collection with respect to patients or siblings compared to controls, as controls were selected by random mailing. We found that the prediction of the high cognitive performer group of patients on a moderate group of siblings was ambiguous and unstable due to low frequency. Generating composite cognition scores might have an impact on finding meaningful trajectory instead of using multivariate cognitive tests. This is one of the major limitations of trajectory modeling (Jones et al., 2001) which dealt with one variable at a time. This study used eight neurocognitive tests but including other tests might lead to different trajectories with different predictions, as the cognitive battery was comprehensive but not complete.

In conclusion, our findings confirmed that cognitive functioning in patients with schizophrenia and their unaffected siblings is heterogeneous. We demonstrated that the cognitive performance of siblings of schizophrenia stayed between that of the patients and the healthy controls. We also identified five distinct cognitive trajectories in patients and four trajectories in siblings, which remained stable during six years follow-up. These trajectories are validated by observing the association with external factors e.g. socio-demographic, clinical and cognitive measures confirming the meaningful cognitive subtypes in patients and siblings. Moreover, cognitive subtypes in patients significantly predicted the sibling subtypes, highlighting the familial contribution to cognition. The study supports neurocognitive profiling as a valuable endophenotype. This profiling approach warrants further evaluation in future molecular studies as well as in studies predicting functional and clinical outcomes.

References

Albus, M., Hubmann, W., Mohr, F., Hecht, S., Hinterberger-Weber, P., Seitz, N.N., Kuchenhoff, H. (2006). Neurocognitive functioning in patients with first-episode schizophrenia : results of a prospective 5-year follow-up study. *European archives of psychiatry and clinical neuroscience* 256:442-451.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders*. Washington DC.

Andruff, H., Carraro, N., Thompson, A., Gaudreau, P., Louvet, B. (2009). Latent Class Growth Modelling: A Tutorial. *Tutorials in Quantitative Methods for Psychology* 5:11-24.

Barder, H.E., Sundet, K., Rund, B.R., Evensen, J., Haahr, U., Ten Velden Hegelstad, W., Joa, I., Johannessen, J.O., Langeveld, H., Larsen, T.K., Melle, I., Opjordsmoen, S., Rossberg, J.I., Simonsen, E., Vaglum, P., McGlashan, T., Friis, S. (2013). Neurocognitive development in first episode psychosis 5 years follow-up: associations between illness severity and cognitive course. *Schizophrenia research* 149:63-69.

Barder, H.E., Sundet, K., Rund, B.R., Evensen, J., Haahr, U., Ten Velden Hegelstad, W., Joa, I., Johannessen, J.O., Langeveld, J., Larsen, T.K., Melle, I., Opjordsmoen, S., Rossberg, J.I., Simonsen, E., Vaglum, P., McGlashan, T., Friis, S. (2013). Ten year neurocognitive trajectories in first-episode psychosis. *Frontiers in human neuroscience* 7:643.

Bilder, R.M., Goldman, R.S., Robinson, D., Reiter, G., Bell, L., Bates, J.A., Pappadopulos, E., Willson, D.F., Alvir, J.M.J., Woerner, M.G., Geisler, S., Kane, J.M., Lieberman, J.A. (2000). Neuropsychology of First-Episode Schizophrenia: Initial Characterization and Clinical Correlates. *AJP* 157:549-559.

Bowie, C.R., Harvey, P.D. (2006). Cognitive deficits and functional outcome in schizophrenia. *Neuropsychiatric Disease and Treatment* 2:531-536.

Bozikas, V.P., Andreou, C. (2011). Longitudinal studies of cognition in first episode psychosis: a systematic review of the literature. *The Australian and New Zealand Journal of Psychiatry* 45:93-108.

Braff, D.L., Freedman, R., Schork, N.J., Gottesman, I.I. (2007). Deconstructing schizophrenia: an overview of the use of endophenotypes in order to understand a complex disorder. *Schizophrenia bulletin* 33:21-32.

Brenner, K., Schmitz, N., Pawliuk, N., Fathalli, F., Joober, R., Ciampi, A., King, S. (2007). Validation of the English and French versions of the Community Assessment of Psychic Experiences (CAPE) with a Montreal community sample. *Schizophrenia research* 95:86-95.

Brzustowicz, L.M., Hodgkinson, K.A., Chow, E.W., Honer, W.G., Bassett, A.S. (2000). Location of a major susceptibility locus for familial schizophrenia on chromosome 1q21-q22. *Science (New York, N.Y.)* 288:678-682.

Brzustowicz, L.M., Simone, J., Mohseni, P., Hayter, J.E., Hodgkinson, K.A., Chow, E.W., Bassett, A.S. (2004). Linkage disequilibrium mapping of schizophrenia susceptibility to the CAPON region of chromosome 1q22. *American Journal of Human Genetics* 74:1057-1063.

Cannon-Spoor, H.E., Potkin, S.G., Wyatt, R.J. (1982). Measurement of premorbid adjustment in chronic schizophrenia. *Schizophrenia bulletin* 8:470-484.

Chen, W.J., Faraone, S.V. (2000). Sustained attention deficits as markers of genetic susceptibility to schizophrenia. *American Journal of Medical Genetics* 97:52-57.

de Rooij, M., Worku, H.M. (2012). A warning concerning the estimation of multinomial logistic models with correlated responses in SAS. *Computer methods and programs in biomedicine* 107:341-346.

Faber, G., Smid, H.G., Van Gool, A.R., Wunderink, L., Wiersma, D., van den Bosch, R.J. (2011). Neurocognition and recovery in first episode psychosis. *Psychiatry research* 188:1-6.

Faraone, S.V., Tsuang, D., Tsuang, M.T. (1999). *Genetics and Mental Disorders: A Guide for Students, Clinicians, and Researchers*. New York: Guilford Press.

Faraone, S.V., Kremen, W.S., Lyons, M.J., Pepple, J.R., Seidman, L.J., Tsuang, M.T. (1995). Diagnostic accuracy and linkage analysis: how useful are schizophrenia spectrum phenotypes?. *The American Journal of Psychiatry* 152:1286-1290.

Formann, A.K. (1984). *Die Latent-Class-Analyse: Einführung in Theorie und Anwendung [Latent class analysis: Introduction to theory and application]*. Germany: Weinheim: Beltz.

Green, M.F., Kern, R.S., Braff, D.L., Mintz, J. (2000). Neurocognitive deficits and functional outcome in schizophrenia: are we measuring the "right stuff"?. *Schizophrenia bulletin* 26:119-136.

Gur, R.E., Nimgaonkar, V.L., Almasy, L., Calkins, M.E., Ragland, J.D., Pogue-Geile, M.F., Kanes, S., Blangero, J., Gur, R.C. (2007). Neurocognitive endophenotypes in a multiplex multigenerational family study of schizophrenia. *The American Journal of Psychiatry* 164:813-819.

Harvey, P.D. (2014). What is the evidence for changes in cognition and functioning over the lifespan in patients with schizophrenia?. *The Journal of clinical psychiatry* 75 Suppl 2:34-38.

Heinrichs, R.W., Zakzanis, K.K. (1998). Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. *Neuropsychology* 12:426-445.

Hilti, C.C., Hilti, L.M., Heinemann, D., Robbins, T., Seifritz, E., Cattapan-Ludewig, K. (2010). Impaired performance on the Rapid Visual Information Processing task (RVIP) could be an endophenotype of schizophrenia. *Psychiatry research* 177:60-64.

Husted, J.A., Lim, S., Chow, E.W.C., Greenwood, C., Bassett, A.S. (2009). Heritability of Neurocognitive Traits in Familial Schizophrenia. *American journal of medical genetics.Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* 150B:845-853.

Irani, F., Kalkstein, S., Moberg, E.A., Moberg, P.J. (2011). Neuropsychological performance in older patients with schizophrenia: a meta-analysis of cross-sectional and longitudinal studies. *Schizophrenia bulletin* 37:1318-1326.

Jones, B.L., Nagin, D.S., Roeder, K. (2001). A SAS Procedure Based on Mixture Models for Estimating Developmental Trajectories. *Sociological Methods & Research* 29:374-393.

Kass, R.E., Raftery, A.E. (1995). Bayes Factors. *Journal of the American Statistical Association* 90:773-795.

Kay, S.R., Fiszbein, A., Opler, L.a. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia bulletin* 13:261-76.

Keefe, R.S., Harvey, P.D. (2012). Cognitive impairment in schizophrenia. *Handbook of Experimental Pharmacology* (213):11-37. doi:11-37.

Kendler, K.S., Diehl, S.R. (1993). The genetics of schizophrenia: a current, genetic-epidemiologic perspective. *Schizophrenia bulletin* 19:261-285.

Keri, S., Janka, Z. (2004). Critical evaluation of cognitive dysfunctions as endophenotypes of schizophrenia. *Acta Psychiatrica Scandinavica* 110:83-91.

Korver, N., Quee, P.J., Boos, H.B.M., Simons, C.J.P., de Haan, L., Investigators, G. (2012). Genetic Risk and Outcome of Psychosis (GROUP), a multi-site longitudinal cohort study focused on gene-environment interaction: objectives, sample characteristics, recruitment and assessment methods. *International Journal of Methods in Psychiatric Research* 21:205-221.

Krabbendam, L., Marcelis, M., Delespaul, P., Jolles, J., van Os, J. (2001). Single or multiple familial cognitive risk factors in schizophrenia?. *American Journal of Medical Genetics* 105:183-188.

Kremen, W.S., Seidman, L.J., Faraone, S.V., Toomey, R., Tsuang, M.T. (2000). The paradox of normal neuropsychological function in schizophrenia. *Journal of abnormal psychology* 109:743-752.

Kremen, W.S., Seidman, L.J., Pepple, J.R., Lyons, M.J., Tsuang, M.T., Faraone, S.V. (1994). Neuropsychological risk indicators for schizophrenia: a review of family studies. *Schizophrenia bulletin* 20:103-119.

Krukow, P., Karakuła-Juchnowicz, H., Juchnowicz, D., Morylowska-Topolska, J., Flis, M., Jonak, K. (2017). Processing speed is associated with differences in IQ and cognitive profiles between patients with schizophrenia and their healthy siblings. *Nordic Journal of Psychiatry* 71:33-41.

Kuss, O., McLerran, D. (2007). A note on the estimation of the multinomial logistic model with correlated responses in SAS. *Computer methods and programs in biomedicine* 87:262-269.

Lancon, C., Auquier, P., Nayt, G., Reine, G. (2000). Stability of the five-factor structure of the Positive and Negative Syndrome Scale (PANSS). *Schizophrenia research* 42:231-239.

Leeson, V.C., Sharma, P., Harrison, M., Ron, M.A., Barnes, T.R., Joyce, E.M. (2011). IQ trajectory, cognitive reserve, and clinical outcome following a first episode of psychosis: a 3-year longitudinal study. *Schizophrenia bulletin* 37:768-777.

Little, R.J.A., Rubin, D.B. (2002). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.

Meijer, J., Simons, C.J., Quee, P.J., Verweij, K., GROUP Investigators. (2012). Cognitive alterations in patients with non-affective psychotic disorder and their unaffected siblings and parents. *Acta Psychiatrica Scandinavica* 125:66-76.

Mohamed, S., Paulsen, J.S., O'Leary, D., Arndt, S., Andreasen, N. (1999). Generalized cognitive deficits in schizophrenia: a study of first-episode patients. *Archives of General Psychiatry* 56:749-754.

Nagin, D.S. (1999). Analyzing Developmental Trajectories: A Semiparametric, Group-Based Approach. *Psychological Methods* 4:139-157.

Nagin, D.S. (2014). Group-based trajectory modeling: an overview. *Annals of Nutrition & Metabolism* 65:205-210.

Nagin, D.S., Odgers, C.L. (2010). Group-based trajectory modeling in clinical research. *Annual review of clinical psychology* 6:109-138.

Niyonkuru, C., Wagner, A.K., Ozawa, H., Amin, K., Goyal, A., Fabio, A. (2013). Group-based trajectory analysis applications for prognostic biomarker model development in severe TBI: a practical example. *Journal of neurotrauma* 30:938-945.

Nuechterlein, K.H., Barch, D.M., Gold, J.M., Goldberg, T.E., Green, M.F., Heaton, R.K. (2004). Identification of separable cognitive factors in schizophrenia. *Schizophrenia research* 72:29-39.

Nuechterlein, K.H., Ventura, J., Subotnik, K.L., Hayata, J.N., Medalia, A., Bell, M.D. (2014). Developing a Cognitive Training Strategy for First-Episode Schizophrenia: Integrating Bottom-Up and Top-Down Approaches. *American journal of psychiatric rehabilitation* 17:225-253.

Palmer, B.W., Heaton, R.K., Paulsen, J.S., Kuck, J., Braff, D., Harris, M.J., Zisook, S., Jeste, D.V. (1997). Is it possible to be schizophrenic yet neuropsychologically normal?. *Neuropsychology* 11:437-446.

Quee, P.J., Alizadeh, B.Z., Aleman, A., van den Heuvel, E.R., GROUP Investigators. (2014). Cognitive subtypes in non-affected siblings of schizophrenia patients: characteristics and profile congruency with affected family members. *Psychological medicine* 44:395-405.

Rodriguez-Sanchez, J.M., Crespo-Facorro, B., Gonzalez-Blanch, C., Perez-Iglesias, R., Alvarez-Jimenez, M., Martinez, O., Vazquez-Barquero, J.L. (2008). Cognitive functioning and negative symptoms in first episode schizophrenia: different patterns of correlates. *Neurotoxicity research* 14:227-235.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* 6:461-464.

Snitz, B.E., Macdonald, A.W.,3rd, Carter, C.S. (2006). Cognitive deficits in unaffected first-degree relatives of schizophrenia patients: a meta-analytic review of putative endophenotypes. *Schizophrenia bulletin* 32:179-194.

Somers, R.H. (1962). A New Asymmetric Measure of Association for Ordinal Variables. *American Sociological Review* 27:799-811.

Szoke, A., Trandafir, A., Dupont, M.E., Meary, A., Schurhoff, F., Leboyer, M. (2008). Longitudinal studies of cognition in schizophrenia: meta-analysis. *The British journal of psychiatry : the journal of mental science* 192:248-257.

Thompson, W.K., Savla, G.N., Vahia, I.V., Depp, C.A., Oâ€™Hara, R., Jeste, D.V., Palmer, B.W. (2012). Characterizing Trajectories of Cognitive Functioning in Older Adults with Schizophrenia: Does Method Matter?. *Schizophrenia research* 143:90-96.

Townsend, L.A., Norman, R.M. (2004). Course of cognitive functioning in first episode schizophrenia spectrum disorders. *Expert review of neurotherapeutics* 4:61-68.

Trandafir, A., Meary, A., Schurhoff, F., Leboyer, M., Szoke, A. (2006). Memory tests in first-degree adult relatives of schizophrenic patients: a meta-analysis. *Schizophrenia research* 81:217-226.

van Os, J., Kenis, G., Rutten, B.P. (2010). The environment and schizophrenia. *Nature* 468:203-212.

Verhage, F. (1964). *Intelligentie en leeftijd: Onderzoek bij Nederlanders van twaalf tot zevenenzeventig jaar*. Assen: Van Gorcum.

Vollema, M.G., Ormel, J. (2000). The reliability of the structured interview for schizotypy-revised. *Schizophrenia bulletin* 26:619-629.

Vollema, M.G., Postma, B. (2002). Neurocognitive correlates of schizotypy in first degree relatives of schizophrenia patients. *Schizophrenia bulletin* 28:367-377.

Wit, E., Heuvel, E.v.d., Romeijn, J. (2012). 'All models are wrong...': an introduction to model uncertainty. *Statistica Neerlandica* 66:217-236.

Supplementary Materials

Supplementary Table S1: Measures of neurocognition

Cognitive domains	Group Tests	Outcome measure
Sustained attention and vigilance	Continuous Performance Test (CPT-HQ) (CPT performance and CPT variance)	An efficiency score [(accuracy/reaction time) ×1000] was created, in which accuracy was measured as the total number of hits (range 0–28) minus the total number of errors (range 0–28), divided by 28. If this calculation of accuracy was non-positive (i.e. the number of errors equaled or exceeded the number of hits), then the accuracy was set equal to 0.005. This score was referred to as 'CPT performance'. Intra-individual variability in reaction time on the CPT was also evaluated (CPT variance), using the standard deviation score of the subject's mean response time on the hit trials (Hilti et al., 2010; Quee et al., 2014).
Verbal learning and memory	Word Learning Task (WLT) (Immediate and Delayed Recall)	Immediate recall (total score of three consecutive trials of 15 words learning) and 'Delayed recall' was assessed after 20 minutes delay.
Global cognitive functioning		
Processing speed	Digit Symbol Substitution	Total raw score (0-133)
Verbal comprehension	Information	Total raw score (0-28)
Working memory	Arithmetic	Total raw score (0-22)
Problem solving and visuospatial abilities	Block Design	Total raw score (0-68)

Supplementary Table S2: Pearson's correlation coefficients of cognitive tests at baseline for patients (n=1,119)

Cognitive measures	CPT performance	CPT variance	Block Design	Digit Symbol	Arith-metic	Information	Immediate Recall	Delayed Recall
CPT variance	-0.54***							
Block Design	0.21***	-0.24***						
Digit Symbol	0.35***	-0.37***	0.43***					
Arithmetic	0.22***	-0.28***	0.52***	0.47***				
Information	0.16***	-0.20***	0.54***	0.43***	0.62***			
Immediate Recall	0.25***	-0.24***	0.32***	0.38***	0.40***	0.39***		
Delayed Recall	0.20***	-0.20***	0.32***	0.34***	0.35***	0.37***	0.77***	
Composite	0.39***	-0.20***	0.70***	0.67***	0.74***	0.75***	0.72***	0 .69***

*** P<0.001

3

Supplementary Table S3: Pearson's correlation coefficients of cognitive tests at baseline for siblings (n=1,059)

Cognitive measures	CPT performance	CPT variance	Block Design	Digit Symbol	Arithme tic	Informat ion	Immediate Recall	Delayed Recall
CPT variance	-0.44***							
Block Design	0.06	-0.15***						
Digit Symbol	0.12***	-0.26***	0.37***					
Arithmetic	0.06	-0.13***	0.50***	0.40***				
Information	0.06	-0.10**	0.44***	0.35***	0.60***			
Immediate Recall	0.11***	-0.15***	0.22***	0.32***	0.30***	0.34***		
Delayed Recall	0.09**	-0.12***	0.20***	0.28***	0.24***	0.29***	0.78***	
Composite	0.25***	-0.08*	0.64***	0.62***	0.72***	0.72***	0.69***	0.65***

*** P<0.001

Supplementary Table S4: Bayesian Information Criterion (BIC) and logged Bayes factor (2*ΔBIC) in patients' and siblings' model

Number of groups	BIC	*ΔBIC*	2**ΔBIC*	Evidence against H_0
Composite cognition (all tests) score of patients (n=1075)				
1	-2400.00			
2	-2042.31	357.69	715.38	
3	-1930.63	111.68	223.36	
4	-1899.23	31.40	62.8	
5	**-1890.56**	**8.67**	**17.34**	**Very strong**
6	-1897.51	-6.95	-13.90	
Composite cognition (all tests) score of siblings (n=1042)				
Number of groups	BIC	*ΔBIC*	2**ΔBIC*	Evidence against H_0
1	-2066.53			
2	-1718.39	348.14	696.28	
3	-1506.76	211.63	423.26	
4	**-1459.45**	**47.31**	**94.62**	**Very strong**
5	-1464.95	-5.50	-11.00	

Table S5: Parameter estimates of trajectory model for patients

Trajectory	Parameter	Estimate	SE	t-value	P-value
Severe	Intercept	-1.57	0.06	-25.78	<0.001
	Linear	0.03	0.03	0.93	0.353
Moderate	Intercept	-1.11	0.11	-10.19	<0.001
	Linear	0.20	0.12	1.61	0.107
	Quadratic	-0.05	0.03	-1.56	0.118
Mild	Intercept	-0.75	0.11	-7.07	<0.001
	Linear	0.39	0.13	3.07	0.002
	Quadratic	-0.10	0.03	-3.18	0.002
Normal	Intercept	0.05	0.05	1.04	0.297
	Linear	-0.01	0.02	-0.32	0.750
High	Intercept	0.71	0.12	6.12	<0.001
	Linear	-0.09	0.04	-2.08	0.038
Drop-out model					
Severe	Drop0	-1.33	2.41	-0.55	0.580
	Drop1	-0.25	0.93	-0.27	0.789
	Drop2	-0.35	1.01	-0.35	0.730
Moderate	Drop0	0.19	1.38	0.14	0.891
	Drop1	0.46	0.88	0.52	0.604
	Drop2	0.65	0.83	0.78	0.436
Mild	Drop0	-0.77	1.19	-0.65	0.518
	Drop1	0.83	2.04	0.40	0.686
	Drop2	0.80	1.35	0.60	0.552
Normal	Drop0	-1.55	0.44	-3.52	0.001
	Drop1	-2.63	1.21	-2.17	0.030
	Drop2	0.90	1.52	0.59	0.555
High	Drop0	1.65	2.56	0.65	0.518
	Drop1	-7.63	5.25	-1.45	0.146
	Drop2	2.40	2.27	1.06	0.291
	Sigma	0.29	0.01	51.79	<0.001
Group	Membership				
Severe	(%)	10.73	1.64	6.56	<0.001
Moderate	(%)	28.44	2.57	11.07	<0.001
Mild	(%)	30.37	2.65	11.46	<0.001
Normal	(%)	26.68	2.38	11.22	<0.001
High	(%)	3.78	1.27	2.98	0.003

BIC= -1902.45 (N=2260), **BIC= -1890.56 (N=1075),** AIC= -1810.88, L= -1778.88; Sigma: The amount of variance in the data accounted for by the model and its significance. Note: Drop0 = constant rate, Drop1 = depends on the previous response (baseline), Drop2 = depends on the two previous responses (baseline and three years).

Table S6: Parameter estimates of trajectory model for siblings

Trajectory	Parameter	Estimate	SE	t-value	P-value
Moderate	Intercept	-1.17	0.04	-26.79	<0.001
	Linear	0.06	0.02	2.82	0.005
Mild	Intercept	-0.53	0.05	-11.61	<0.001
	Linear	0.03	0.02	1.51	0.131
Normal	Intercept	-0.33	0.08	-4.17	<0.001
	Linear	0.26	0.09	2.91	0.004
	Quadratic	-0.06	0.02	-2.49	0.013
High	Intercept	0.43	0.02	14.11	<0.001
	Linear	-0.003	0.01	-0.23	0.818
Drop-out model					
Moderate	Drop0	-1.10	1.29	-0.86	0.390
	Drop1	-0.08	0.86	-0.09	0.928
	Drop2	-0.28	1.08	-0.26	0.793
Mild	Drop0	-0.22	1.55	-0.14	0.886
	Drop1	2.33	2.01	1.16	0.248
	Drop2	0.52	1.61	0.32	0.746
Normal	Drop0	-2.41	0.50	-4.85	<0.001
	Drop1	-4.42	1.97	-2.24	0.025
	Drop2	-0.23	1.38	-0.17	0.869
High	Drop0	-0.45	0.74	-0.61	0.544
	Drop1	-1.14	1.25	-0.92	0.359
	Drop2	-2.11	1.12	-1.89	0.059
	Sigma	0.26	0.01	55.67	<0.001
Group	Membership				
Moderate	(%)	13.03	1.35	9.64	<0.001
Mild	(%)	25.13	2.84	8.86	<0.001
Normal	(%)	37.61	2.83	13.30	<0.001
High	(%)	24.23	1.81	13.35	<0.001

BIC= -1469.44 (N=2316), **BIC= -1459.45 (N=1042),** AIC= -1397.59, L= -1372.59; Sigma: The amount of variance in the data accounted for by the model and its significance. Note: Drop0 = constant rate, Drop1 = depends on the previous response (baseline), Drop2 = depends on the two previous responses (baseline and three years).

Table S7: Contingency table of subtypes for the pair of siblings and patients*

Patient Subtypes	Sibling Subtypes				Total	Somers' *D*±ASE
	High	Normal	Mild	Moderate		
High	15 (6.7)	9 (10.9)	4 (6.8)	0 (3.5)	28	
Normal	119 (72.0)	114 (117.6)	56 (73.4)	12 (38.0)	301	
Mild	68 (78.7)	155 (128.5)	70 (80.3)	36 (41.5)	329	0.29±0.02
Moderate	47 (76.1)	112 (124.2)	98 (77.6)	61 (40.1)	318	
Severe	7 (22.5)	28 (36.7)	33 (22.9)	26 (11.9)	94	
Total	256	418	261	135	1070	

*Table represents Observed (Expected) frequencies and Somers' *D* statistic with asymptotic standard error (ASE)

CHAPTER 4

Long-term course of negative symptom subdomains and relationship with outcome in patients with a psychotic disorder

Annemarie P. M. Stiekema
Md. Atiqul Islam
Edith J. Liemburg
Stynke Castelein
Edwin R. van den Heuvel
Jaap van Weeghel
André Aleman
Lisette van der Meer
GROUP-investigators

In revision (Schizophrenia Research)

Abstract

Background: Negative symptoms can be divided into social amotivation (SA) and expressive deficits (ED). We investigated *i*) the course of SA and ED over six years, *ii*) whether SA and ED were related to functioning and quality of life six years later, *iii*) whether subgroups based on the course of SA and ED could be identified, and *iv*) the relationship between subgroups and outcomes over six years.

Methods: Measurements at baseline, three and six years from 1067 patients participating in the Genetic Risk and Outcome of Psychosis (GROUP) project were used. We applied mixed models analysis, regression analysis and trajectory analyses.

Results: Results showed that SA and ED decreased over time. Lower symptom levels were related to better functioning (SA, ED) and quality of life (SA) at six years. Within each subdomain, four subgroups were identified: a steady low course (±60%), a course where symptoms increased (±15%), a subgroup where symptoms decreased (±15%), within SA a decreased course but starting at a higher level (±6%) and within ED a course with relatively stable high ED scores (±6%). In general, the low SA and ED groups showed better outcomes than the other subgroups within each domain.

Conclusion: A substantial number of patients do not follow a stable course of negative symptoms. This heterogeneity is related to outcomes later in life. When evaluating treatments, the effects in subgroups with fluctuating symptom levels may be averaged out by the larger groups showing steady low negative symptom levels.

Keywords: negative symptoms; social amotivation; expressive deficits; functional outcome; long-term course

1. Introduction

Psychotic disorders such as schizophrenia are characterized by a variable presentation of positive symptoms, negative symptoms, and cognitive deficits (Mueser and McGurk, 2004; American Psychiatric Association, 2000). Although positive symptoms are usually most dominant in the acute phase of illness, negative symptoms are considered to be most disabling, due to their persistent nature and profound relationship with poor functional outcomes (Bobes et al., 2010; Ventura et al., 2009). Despite the growing body of research, negative symptoms still are difficult to treat; both pharmacological and psychosocial interventions only have a limited effect, if any at all (Savill et al., 2014). The development of treatments for negative symptoms is difficult due to their heterogeneous nature. Current research therefore aims to diminish this heterogeneity by grouping negative symptoms into two subdomains: social amotivation (SA) and expressive deficits (ED) (Messinger et al., 2011; Foussias et al., 2014).

SA encompasses social and emotional withdrawal and reflects diminished interest in or affective commitment to the social environment. SA is thought to be the result of a deficit in anticipating on pleasurable events and activities (Foussias et al., 2014; Buck and Lysaker, 2013). ED involves blunted affect, poverty of speech, and motor retardation. ED reflects a diminished expressive responsiveness that is observed in verbal and non-verbal communication, which is thought to be caused by, or least related to, neurocognitive deficits (Liemburg et al., 2013; Bell et al., 2013; Ergul and UCok, 2015; Hartmann-Riemer et al., 2015). There is ample evidence for a strong relationship between SA and global functioning (Foussias et al., 2011; Fervaha et al., 2014; Rocca et al., 2014). The associations of ED with functioning were found to be less strong (Foussias et al., 2011; Strauss et al., 2013). However, we recently reported that ED, but not SA, predicted residential living status in a chronic population with psychotic disorders (Stiekema et al., 2016), indicating that ED may in fact be related to daily functioning. However, the extent to which scores on subdomains are consistent over time remains unclear. The few studies that have investigated the longitudinal course of SA and ED showed mixed results, varying from long-term stability of both domains (Galderisi et al., 2013), of ED but not of SA (Ergul and UCok, 2015), and vice versa (Norman et al., 2015).

In the current study, we first investigated whether baseline levels of SA and ED were related to functioning (global functioning, social functioning, independent living, and engagement in work or study) and quality of life six years later. Secondly, we examined whether subgroups with different longitudinal courses of SA and ED could be identified. Finally, we tested whether changes in subdomain scores were differentially associated with changes in functioning and quality of life. In accordance with our previous findings, we hypothesized that both subdomains would be related to global functioning and engagement in work or study, that SA would be most strongly related to social functioning and quality of life, while ED would be related to non-independent living status.

2. Methods

2.1. Study design

We used data from the Genetic Risk and Outcome of Psychosis (GROUP) project, in which outpatients and inpatients with a psychotic disorder between 16 and 50 years were recruited from 36 sites in the Netherlands. The procedure of recruitment, informed consent, approval by the accredited Medical Ethics Review Committee (METC), assessment and population characteristics have been described in

detail elsewhere (Korver et al., 2012). Between April 2004 and December 2013, participants were assessed at baseline and three and six years thereafter.

2.2. Participants

The GROUP sample consisted of 1119 patients and 586 healthy controls at baseline (Korver et al., 2012). Fifty-three patients were excluded because their diagnosis was missing (n = 4), unclear (n=21) or other than primary psychotic (n = 27), so that 1067 patients were included in the analysis.

2.3. Assessment

Psychopathology was assessed with the Positive and Negative Syndrome Scale (PANSS (Kay et al., 1987)). SA score was calculated as the sum of N2 (emotional withdrawal), N4 (passive/apathetic social withdrawal), and G16 (active social avoidance). ED was the sum of N1 (flat affect), N3 (poor rapport), N6 (lack of spontaneity), G5 (mannerisms and posturing), G7 (motor retardation), and G13 (avolition) (Liemburg et al., 2013).

Global functioning was measured with the Global Assessment of Functioning Disability scale (GAF-D (American Psychiatric Association, 2000)), on an anchored scale from 1 (most severe) to 100 (excellent functioning).

Social functioning was measured with the Social Functioning Scale (SFS) (Birchwood et al., 1990), filled out by the participant at three and six years. The SFS score was computed as the mean of the seven subscales scaled scores.

Independent living (single or with partner or own family vs. living with parents or other family members or sheltered living) and engagement in work or study (work/study vs. no work/study) were also used as functional outcome measures.

Quality of life was assessed with the World Health Organization Quality of Life-BREF (WHO-QOL-BREF (Trompenaars et al., 2005)), including four domains of quality of life: physical health, psychological, social relationships, and environment.

Neurocognition was based on a composite score (mean z-scores) of the Continuous Performance Test, Word Learning Task immediate recall and delayed recall and recognition, and WAIS-III Symbol Substitution, Information, Arithmetic and Block Design. Healthy control subjects were used to obtain age and gender specific z-scores for patients.

2.4. Statistical analysis

2.4.1. Longitudinal course and subgroups of SA and ED

Baseline characteristics of completers versus non-completers (patients who did not participate in the three and/or six-year measurement) were compared using the Kruskal-Wallis test for continuous variables and Chi-square tests for categorical variables.

We examined the change in overall SA and ED over time. We checked the average plot on original (non-imputed) data and conducted a linear mixed model was on the imputed scores for SA and ED separately, including only the fixed effect of time as a categorical independent variable.

Group-based trajectory modeling (Jones et al., 2001; Niyonkuru et al., 2013) was conducted using PROC TRAJ procedure in order to identify clusters of patients following similar patterns within each subgroup (i.e. separately on original scores for SA and ED) over time. Since SA and ED are

continuous, a normal distribution model was specified by identifying a minimum and maximum value outside the range of observed SA or ED values. A first order linear and second order quadratic polynomial model was fitted to determine the number of groups over time. A maximum likelihood method was applied to estimate parameters, including group sizes and shapes of trajectories including patients with missing data. The Bayesian Information Criterion (BIC) and logged Bayes factor (2*ΔBIC) were used to select the number of clusters or subgroups that best fit the data (Jones et al., 2001).

Differences between the identified subgroups in baseline demographic and clinical characteristics were examined using the Kruskal-Wallis test for continuous variables and Chi-square or Fishers exact tests for categorical variables. Pairwise comparisons were corrected for multiple testing using Bonferroni correction.

2.4.2. Multiple imputation

Multiple imputation was applied to address missing data in outcomes and independent variables. Since the data showed arbitrary missing patterns, a fully conditional specification (FCS) predicted mean matching (PMM) method was used to impute missing values for both continuous and categorical variables (van Buuren, 2007), which generated 10 imputed datasets of 1067 patients. All variables (demographic, clinical and neurocognitive as well as the outcomes) were included in the imputation model at three time points. Subsequently, the analyses as described below were performed on each dataset with our different proposed models and parameter estimates were pooled with Rubin's rule (Rubin, 1987).

2.4.3. Predicting outcome at six years

Multiple linear regression analysis was conducted on six year imputed GAF, SFS and WHO-QOL scores as well as logistic regressions on living situation and work activities to investigate the relationship with baseline SA and ED. Baseline SA and ED were entered into the first block, the confounders gender, duration of illness, positive symptoms (PANSS positive subscale), and neurocognition (composite score) into the second block.

2.4.4. Subgroups predicting functioning

Since repeated measures within patients are correlated, linear mixed models were performed on continuous outcomes and generalized linear mixed models were conducted on categorical outcomes, examining whether changes in subgroups within each subdomain were associated with changes in functioning and quality of life over time. A random intercept (patients) mixed model was chosen. For continuous outcomes, the parameter estimates and their variance components were estimated with restricted maximum likelihood (REML). An adaptive Gaussian quadrature with 10 quadrature points was used to estimate the parameters and their associated standard errors for binary outcomes. The independent variables (including subgroups of subdomains and time as a categorical measure) in the statistical model were all considered as fixed effects. The interaction of the categorical subgroups of subdomains with time was one of our interests. The first block of variables in the model contained the subgroups of subdomains and the second block controlled for gender, duration of illness, positive symptoms, and neurocognition).

Pooled Type-III tests of fixed effects p-values (Rubin, 1987; Li et al., 1991) were used to conclude the marginal effects on different outcomes. Additionally, the mean differences of SA or ED subgroups were compared and corrected for multiple testing using Bonferroni correction. All analyses were done using two-tailed tests at α=0.05. Analyses were performed using Statistical Analysis System (SAS), version 9.4 (SAS Institute, 2013).

3. Results

Baseline characteristics are shown in Table 1. Compared completers, non-completers on average had a significantly shorter duration of illness (3.78 vs 4.53 years), lower education levels (3.75 vs 4.19) and higher SA scores (6.45 vs 6.00) and ED scores (11.34 vs 10.43).

Table 1: Baseline demographic and clinical characteristics of participants (n=1067).

	N	Mean (standard deviation) or percentage
Demographics		
Age, years	1059	27.1 (7.24)
Gender, male	1067	77.1 %
Education[a]	1015	4.0 (2.06)
Caucasian	823	79.2 %
Marital status	1051	
Not married	929	88.4 %
Married/living together	93	8.8 %
Divorced/widowhood	29	2.8 %
Residential status	991	
Single or with partner/family	433	43.7 %
With parent(s) or sheltered living	494	49.8 %
Other	64	6.5 %
Clinical characteristics		
Diagnosis	1067	
Schizophrenia	722	67.7 %
Schizo-affective disorder	120	11.2 %
Psychosis NOS	113	10.6 %
Schizophreniform	62	5.8 %
Other[b]	50	4.7 %
Duration of illness, years	1011	4.2 (3.83)
Recent onset psychosis[c]	1067	32.6 %
Number of hospitalizations	895	1.9
Number of psychotic episodes	1041	1.7
GAF	970	54.4 (16.03)
SFS total[e]	-	-
PANSS total	1014	54.9 (16.77)
PANSS positive	1015	12.7 (5.33)
PANSS negative	1012	14.1 (6.01)
PANSS general	1014	28.1 (8.40)
PANSS social amotivation	1001	6.2 (3.09)
PANSS expressive deficits	996	10,79 (4.76)
WHO-QOL total	946	88.4 (14.82)

GAF: Global Assessment of Functioning; SFS: Social Functioning Scale; PANSS: Positive and Negative Syndrome Scale; WHO-QoL: World Health Organization Quality of Life. [a]Education (Verhage): range 1 (no education), 2 (education but no diploma), 3–5 (school diploma) to 8 (university degree); [b]32 Brief psychotic disorder (2.9 %), 22 delusional disorder (2.1 %), 1 psychotic disorder due to medical condition (0.1%); [c]First psychotic episode <2 years prior to baseline measurement; [d]Dose equivalents of chlorpromazine were evaluated using the methods of (Gardner et al. 2010); [e]The SFS was only administered at the 3 and 6 year measurements.

3.1. Missing data

See Supplementary Table 1 for an overview of missing data. All analyses were conducted on imputed data, except for the trajectory analyses.

3.2. Longitudinal course of SA and ED

SA and ED both significantly reduced over time (overall pooled Type-III fixed effect $F_{2, 2120} = 65.69$, $p < .001$; $F_{2, 2120} = 84.90$, $p < .001$) (See the overall profile in Figure 1).

3.3. Predicting outcome at six years

Lower baseline SA predicted a higher level of global functioning (GAF; β = -0.73, t = -3.46, $p = .001$), social functioning (SFS; β = -0.70, t = -6.30, p <0 .001), better quality of life (β = -0.64, t = -3.62, p <.001) and engagement in work or study (β = -0.08, t = -2.18, $p = 0.03$) six years later.

Lower baseline ED predicted a higher level of global functioning (GAF; β = -0.36, t = -2.09, p = .04) and social functioning (SFS; β = -0.32, t = -3.40, p = 0 .002) six years later.

3.4. Longitudinal course of subgroups of SA and ED

Within each subdomain, four subgroups with a different course of negative symptoms could be identified (Supplementary Table 2). The patterns were similar within each subdomain: low, decreased (-low), (decreased-) high and increased. Figure 1 shows the course of each subgroup and the percentage of patients following each course. Demographic differences between the subgroups can be found in Table 2.

3.5. Relationship between subgroups and the level of functioning

We have examined differences between the subgroups within each subdomain with regard to the level of outcomes at each time point. The low SA group scored higher (better) than the other groups on the GAF, SFS and the WHOQOL-BREF at all time points. The low ED group scored higher (better) than the other groups at all time points on the GAF and SFS, except for the six-year measurement compared to the decreased ED. These differences and other significant differences between specific groups are presented in Supplementary Table 3 and graphically represented in Figure 2 and Supplementary Figure 1.

3.6. Relationship between subgroups and the course of functioning

Significant differences in the course of outcomes were found between the subgroups within SA and within ED. Findings are presented in Supplementary Table 4 and Table 3 and graphically presented in Figure 2 (global functioning, social functioning and quality of life) and Supplementary Figure 1 (living situation and engagement in work/study).

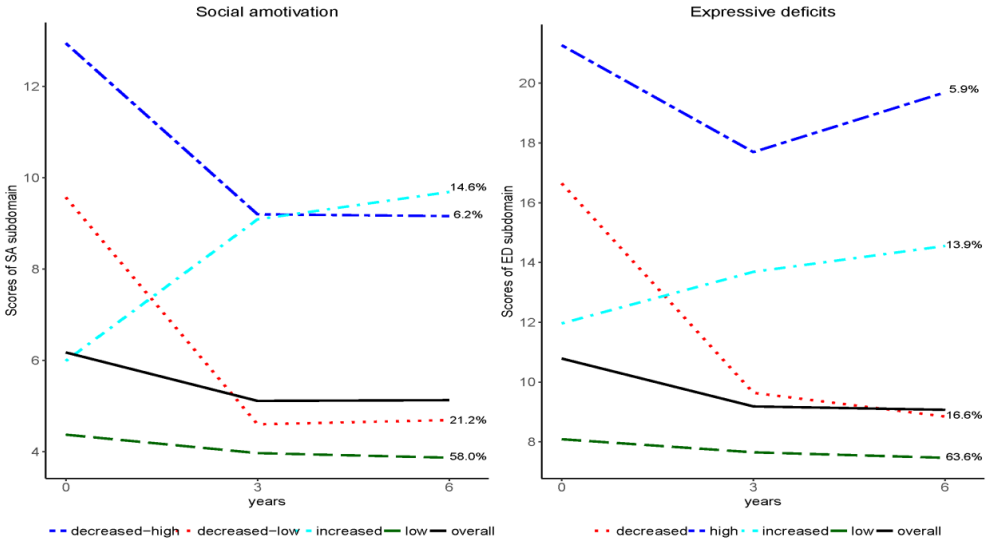


Figure 1: Subgroups with a different course of symptoms over a period of 6 years within SA and within ED.

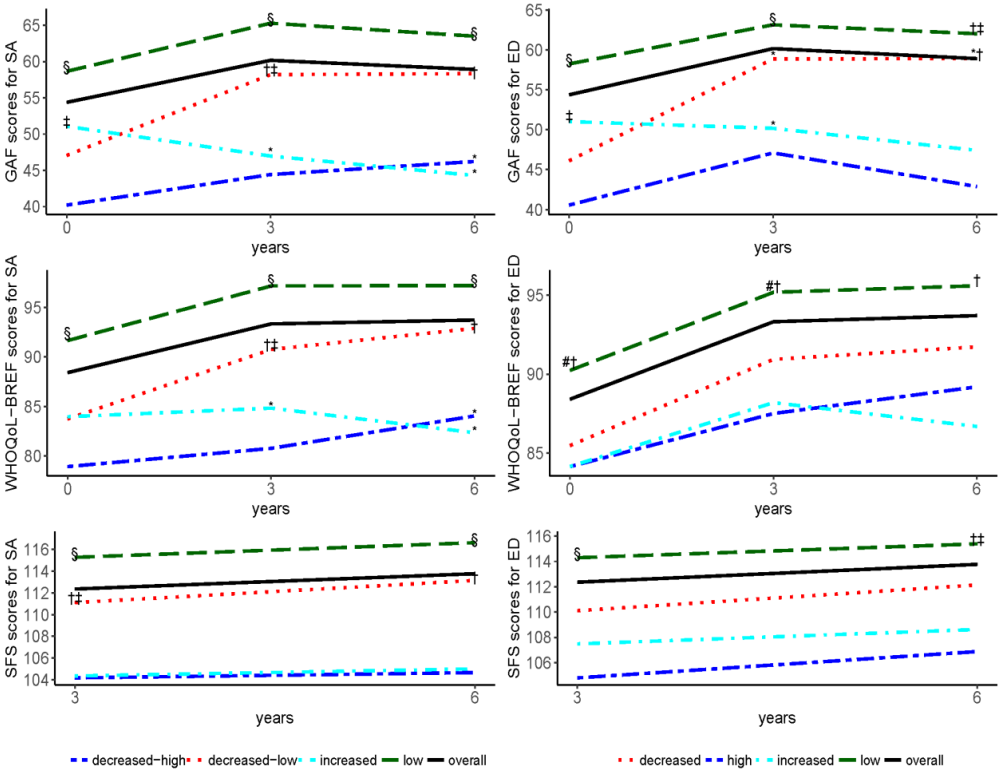


Figure 2: Average values of global and social functioning scores and quality of life (based on imputed data). Analyses were controlled for gender, duration of illness, positive symptoms and cognition. Bonferroni correction (p-value multiplied by 6) was used for the pairwise comparisons. Significant differences between two groups are indicated only at the upper group. Indicated are *significantly different *course* compared to the low group, between study entry and the marked time point (mixed models), significantly different *level* compared to §all other subgroups, #the decreased (-low) group, †the increased group and ‡the (decreased-) high group at the marked time point (pairwise comparisons).

Table 2: Baseline demographic and clinical characteristics per subgroup[a]

	SA				ED			
	Low (n=670)	Decreased-low (n=120)	Increased (n=223)	Decreased-high (n=54)	Low (n=715)	Decreased (n=180)	Increased (n=114)	High (n=58)
Demographic characteristics								
Age, years	26.9 (7.2)#	26.8 (7.2)	28.4 (7.5)	28.4 (7.0)	27.4#	25.4 (6.4)†	28.5 (7.1)‡	26.0 (7.7)
Gender, male	73.3‡	82.1	81.7	94.4	75.0‡	82.2	76.3	89.7
Education	4.2 (2.0)‡	3.8 (2.1)	3.7 (2.0)	3.4 (2.0)	4.2 (2.0)*	3.7 (2.0)	3.7 (2.0)	3.4 (2.1)
Caucasian	82.1†	77.4	70.1	71.7	80.3‡	80.1	77.5	66.1
Marital status								
Not married	86.6	90.0	92.4	94.4	86.6#	82.2	87.6	94.8
Married/living together	10.0	8.1	6.8	1.9	10.0	17.8	10.6	3.4
Divorced/widowhood	3.3	1.8	0.8	3.7	3.4	1.1	1.8	1.7
Residential status								
Single	33.9	32.4	34.5	34.6	36.4	27.5	32.1	24.1
With parent(s)	39.7	39.1	40.9	42.3	37.7	43.9	42.5	48.1
With partner/family	11.7	8.2	7.3	1.9	12.0#	5.3	8.5	3.7
Sheltered living	8.4	13.0	11.8	13.5	8.6	11.7	12.3	16.7
Other	6.3	7.2	5.5	7.7	5.3#	11.7	4.7	7.4
Clinical characteristics								
Diagnosis								
Schizophrenia	61.6*	74.9	80.8	83.3	62.8*	73.9	79.8	84.5
Schizo-affective disorder	12.5	10.3	5.8	11.1	13.0	7.8	7.9	6.9
Psychosis NOS	12.5‡	10.3	5.0	-	11.7	10.6	6.1	5.2
Schizophreniform	7.0#	2.2	5.8	5.6	6.0	5.6	6.1	3.4
Other	6.3	2.3	2.5	-	6.5†	2.1	-	-
Duration of illness, years	4.2 (3.8)	3.9 (3.2)	4.7 (4.6)	5.1 (4.3)	4.3	3.8 (3.4)	4.5 (3.9)	4.1 (3.8)
Recent onset psychosis	32.5	31.8	35.8	29.6	31.9	36.1	33.3	29.3
Number of hospitalizations	1.7 (1.9)	2.1 (2.7)	2.2 (3.0)	2.0 (2.2)	1.8	2.2 (3.2)	1.8 (1.5)	1.9 (2.3)
Number of psychotic episodes	1.8 (1.2)	1.63 (1.1)	1.6 (1.0)	1.8 (1.3)	1.8	1.7 (1.1)	1.6 (.9)	1.7 (1.2)
Chlorpromazine equivalent	322.8 (283.3)	359.9 (342.0)	356.7 (275.9)	342.2 (280.7)	313.3 (291.6)‡	383.2 (326.6)	356.9 (278.2)	420.4 (241.6)

Table 2: Baseline demographic and clinical characteristics per subgroup[a]-continued

SA scores	4.4 (1.5)*	9.6 (1.6)*	6.0 (1.8)*	12.9 (2.3)*	5.2 (2.6)*	8.5 (2.7)†	6.7 (2.7)*	10.0 (3.6)
ED scores	9.1 (3.6)*	13.7 (4.8)*	11.6 (4.3)	16.7 (5.9)	8.1 (2.1)*	16.6 (2.4)†	12.0 (2.7)*	21.3 (3.4)
GAF	58.7 (16.4)*	47.1 (12.2)‡	51.1 (13.0)‡	40.4 (9.6)*	58.2 (15.7)*	46.1 (13.2)	51.0 (14.3)‡	40.6 (12.4)
PANSS total	47.8 (13.0)*	67.5 (14.9)*	57.6 (12.7)*	78.6 (16.9)*	48.6 (13.6)*	68.4 (14.1)†	58.5 (12.4)‡	79.1 (16.7)
PANSS positive	11.6 (4.6)*	14.8 (6.1)	13.3 (4.9)	16.3 (6.5)	12.1 (5.2)*	14.0 (5.7)	13.4 (4.7)	15.5 (5.7)
PANSS negative	11.2 (4.1)*	19.2 (5.0)*	14.9 (4.7)*	24.0 (5.8)*	11.1 (3.8)*	20.5 (4.0)†	15.7 (4.1)‡	25.2 (5.7)
PANSS general	25.0 (6.9)*	33.47 (8.1)†	29.4 (6.7)*	38.2 (9.1)	25.4 (7.0)*	33.9 (8.3)†	29.5 (6.6)‡	38.4 (9.1)
WHO-QOL-BREF	91.6 (14.5)*	83.7 (13.9)	84.0 (13.0)	78.9 (14.5)	90.2 (15.2)*	85.5 (13.5)	84.1 (12.6)	84.1 (15.0)
Neurocognition	-.50 (.6)†	-.64 (.6)	-.55 (.7)	-.65 (.7)	-.47 (.6)*	-.74 (.7)	-.67 (.7)	-.73 (.6)

[a] Differences between the subgroups were tested using the Kruskal-Wallis test for continuous variables, the Chi-square test for categorical variables and the Fishers exact test when expected counts were less than five. P-values were multiplied by the number of comparisons (six) to correct for inflated experimentwise error. For the sake of clarity, significant differences between two groups are indicated only in the first column of the groups. * Statistically significant difference compared to all other groups within the subdomain at $\alpha = .05$; [#] Statistically significant difference compared to decreased(-low) group within the subdomain at $\alpha = .05$; † Statistically significant difference compared to increased group within the subdomain at $\alpha = .05$; ‡ Statistically significant difference compared to (decreased-)high group within the subdomain at $\alpha = .05$.

Table 3: Pooled parameter estimates of mixed models analyses adjusted for gender, duration of illness. neurocognition and positive symptoms*.

	Living situation				Work/study				GAF			
	B	SE	95% CI	p-value	B	SE	95% CI	p-value	B	SE	95% CI	p-value
Social amotivation (SA)												
Intercept	0.37	0.32	-0.27; 1.01	0.253	3.25	0.27	2.71; 3.79	<.001	74.28	1.04	72.24; 76.32	<.001
Decreased-high SA	-0.57	0.61	-1.76; 0.62	0.351	-0.92	0.40	-1.71; -0.14	0.021	-11.92	1.97	-15.78; -8.07	<.001
Decreased-low SA	-0.04	0.33	-0.68; .60	0.901	-0.16	0.26	-0.66; 0.35	0.545	-7.24	1.09	-9.38; -5.10	<.001
Increased SA	-0.21	0.42	-1.04; 0.62	0.620	0.06	0.35	-0.63; 0.75	0.864	-5.34	1.38	-8.04; -2.64	<.001
Time: 3 years	0.98	0.20	0.59; 1.37	<.001	0.11	0.23	-0.36; 0.58	0.644	4.03	0.73	2.60;5.48	<.001
Time: 6 years	2.11	0.30	1.50; 2.72	<.001	0.23	0.20	-0.17; 0.62	0.256	2.80	0.72	1.38;4.22	<.001
Decreased-high SA*3 years	-1.04	0.68	-2.37; 0.30	0.128	-0.40	0.56	-1.51; 0.72	0.480	-0.58	2.53	-5.58; 4.43	0.821
Decreased-high SA*6 years	-0.49	0.98	-2.50; 1.51	0.617	-0.21	0.57	-1.34; 0.91	0.708	2.19	2.90	-3.63; 8.02	0.453
Decreased-low SA*3 years	-0.73	0.35	-1.41; -0.04	0.039	-0.37	0.43	-1.23; 0.49	0.388	1.39	1.43	-1.44; 4.21	0.332
Decreased-low SA*6 years	-0.48	0.44	-1.36; 0.40	0.280	-0.24	0.36	-0.95; 0.47	0.509	3.12	1.37	0.41; 5.82	0.024
Increased SA*3 years	-0.09	0.46	-0.98; 0.81	0.851	-1.11	0.44	-1.97; -0.26	0.011	-7.42	1.66	-10.68; -4.17	<.001
Increased SA*6 years	-0.63	0.53	-1.68; 0.41	0.231	-1.33	0.42	-2.16; -0.50	0.002	-5.65	1.79	-9.20; -2.10	0.002
Expressive deficits (ED)	B	SE	95% CI	p-value	B	SE	95% CI	p-value	B	SE	95% CI	p-value
Intercept	0.52	0.32	-0.11; 1.15	0.107	3.40	0.28	2.86; 3.95	<.001	75.34	1.07	73.24; 77.43	<.001
High ED	-1.17	0.59	-2.32; -0.03	0.045	-0.60	0.41	-1.42; 0.21	0.147	-11.66	1.90	-15.38; -7.93	<.001
Decreased ED	-0.94	0.35	-1.64; -0.25	0.008	-0.14	0.28	-0.69; 0.41	0.616	-8.46	1.19	-10.80; -6.11	<.001
Increased ED	-0.54	0.43	-1.38; 0.30	0.204	0.05	0.34	-0.61; 0.71	0.883	-4.85	1.43	-7.64; -2.05	<.001
Time: 3 years	0.85	0.18	0.49; 1.21	<.001	-0.03	0.21	-0.45; 0.39	0.896	2.89	0.68	1.55; 4.22	<.001
Time: 6 years	1.88	0.28	1.29; 2.47	<.001	0.02	0.21	-0.41; 0.45	0.913	1.90	0.78	0.35; 3.45	0.018
High ED*3 years	-0.90	0.67	-2.23; 0.42	0.179	-0.15	0.64	-1.43; 1.14	0.820	1.99	2.58	-3.14; 7.11	0.443
High ED*6 years	-0.53	0.94	-2.48; 1.42	0.579	-0.34	0.53	-1.38; 0.69	0.515	3.14	2.91	-2.73; 9.01	0.287
Decreased ED*3 years	0.14	0.38	-0.60; 0.88	0.708	-0.09	0.39	-0.86; 0.67	0.811	4.52	1.44	1.69; 7.36	0.002
Decreased ED*6 years	0.56	0.42	-0.26; 1.38	0.178	0.05	0.38	-0.70; 0.81	0.895	6.37	1.85	2.62; 10.12	0.002
Increased ED*3 years	-0.50	0.46	-1.41; 0.40	0.276	-0.94	0.42	-1.77; -0.11	0.027	-3.62	1.74	-7.05; -0.19	0.039
Increased ED*6 years	-0.37	0.48	-1.32; 0.57	0.439	-0.42	0.51	-1.426; 0.59	0.413	-2.51	2.02	-6.55; 1.52	0.217

Table 3: Pooled parameter estimates of mixed models analyses adjusted for gender, duration of illness. neurocognition and positive symptoms*-continued

	SFS				WHOQOL-BREF			
	B	SE	95% CI	p-value	B	SE	95% CI	p-value
Social amotivation (SA)								
Intercept	119.75	0.82	118.12; 121.38	<.001	98.03	1.22	95.61; 100.46	<.001
Decreased-high SA	-8.51	1.33	-11.14; -5.89	<.001	-9.59	2.05	-13.61; -5.57	<.001
Decreased-low SA	-3.69	0.77	-5.22; -2.16	<.001	-6.04	1.12	-8.24; -3.84	<.001
Increased SA	-8.97	0.91	-10.76; -7.19	<.001	-6.54	1.43	-9.34; -3.74	<.001
Time: 3 years	-	-	-		4.09	0.64	2.81; 5.37	<.001
Time: 6 years	1.30	0.33	0.64; 1.96	<.001	3.79	0.66	2.47; 5.11	<.001
Decreased-high SA*3 years	-	-	-	-	-2.33	2.13	-6.53; 1.87	0.276
Decreased-high SA*6 years	1.15	1.47	-1.86; 4.16	0.440	1.79	2.24	-2.64; 6.21	0.426
Decreased-low SA*3 years	-	-	-	-	0.57	1.20	-1.80; 2.94	0.635
Decreased-low SA*6 years	0.42	0.72	-1.02; 1.86	0.563	2.67	1.17	0.37; 4.97	0.023
Increased SA*3 years	-	-	-	-	-3.49	1.47	-6.39; -0.60	0.018
Increased SA*6 years	0.24	0.75	-1.23; 1.71	0.748	-4.12	1.46	-6.99; -1.25	0.005
	B	SE	95% CI	p-value	B	SE	95% CI	p-value
Expressive deficits (ED)								
Intercept	120.06	0.82	118.42; 121.69	<.001	98.38	1.25	95.92; 100.85	<.001
High ED	-7.05	1.40	-9.84; -4.25	<.001	-3.52	2.04	-7.53; 0.48	0.085
Decreased ED	-3.37	0.77	-4.89; -1.84	<.001	-3.35	1.21	-5.73; -0.97	0.006
Increased ED	-4.92	0.96	-6.81; -3.02	<.001	-4.65	1.44	-7.47; -1.83	0.001
Time: 3 years	-	-	-	-	3.72	0.61	2.51; 4.93	<.001
Time: 6 years	1.13	0.33	0.48; 1.77	<.001	3.58	0.63	2.32; 4.85	<.001
High ED*3 years	-	-	-	-	-1.76	2.16	-6.03; 2.52	0.418
High ED*6 years	1.67	1.53	-1.46; 4.84	0.280	1.04	2.61	-4.25; 6.3	0.693
Decreased ED*3 years	-	-	-	-	-0.05	1.34	-2.71; 2.61	0.970
Decreased ED*6 years	1.07	0.77	-0.48; 2.26	0.172	1.73	1.35	-0.94; 4.41	0.202
Increased ED*3 years	-	-	-	-	-0.35	1.49	-3.27;2.58	0.817
Increased ED*6 years	1.05	0.87	-0.68; 2.78	0.230	-0.12	1.65	-3.40; 3.16	0.941

Reference category for SA is the low SA group, reference category for ED is low ED group. Reference category for time is baseline expect for the SFS. where the three-year measurement was the reference category in the absence of a baseline measurement.

4. Discussion

The aim of this study was fourfold. Firstly, we examined the course of SA and ED over six years. Secondly, we investigated whether SA and ED at baseline were related to functioning and quality of life six years later. Thirdly, we examined whether we could disentangle the heterogeneity of negative symptoms by classifying patients into different subgroups based on the course of SA and ED over time. And finally, we investigated to what extent these subgroups differed in their level and course of functioning over six years. We distinguished separate groups within the subdomains SA and ED, following a different course of negative symptoms over time. Furthermore, we demonstrated that the course of negative symptomatology over time was related to the level and courses of functioning and quality of life over a period of six years.

4.1. Longitudinal course and subgroups of SA and ED

In contrast with previous findings that suggest a stable course of negative symptoms in SA and ED across five years (Galderisi et al., 2013), the current findings demonstrate that whilst this is the case for approximately two third of the patients, approximately one third follows a less stable course. This one third is divided into subgroups showing a decreased or increased SA and ED course over time. The identification of a stably high ED group, but not a stably high SA group, may support previous suggestions that ED is more persistent (Liemburg et al., 2013; Ergul and UCok, 2015), although only a small proportion of patients seems to suffer from stably high ED (6%).

According to the literature, improvement of negative symptoms often takes place in the first few years of illness (Evensen et al., 2012; Eaton et al., 1995; Hovington et al., 2012) and an increase in negative symptoms is predominantly found in chronic patients (Chang et al., 2011). By introducing subgroups, we were able to demonstrate a more detailed account of negative symptom development, namely a differential course for subgroups of patients for both SA and ED over time. Both the decrease and increase of negative symptoms took place mainly in the first three years of the study, indicating that the variability in symptom level may diminish with a longer duration of illness and it could reflect fluctuations in patients within their first years of illness (subgroups did not differ with regard to duration of illness).

In sum, these results show that, not only do the subdomains SA and ED provide more information with regard to the heterogeneity in symptom presentation, the course of negative symptom development over time seems to be more variable between subgroups of patients than was previously assumed, when looking within each subdomain.

4.2. Associations with outcome

The subgroups within SA and ED were differentially related to the level and course of functioning and quality of life across six years. The lower symptom groups generally showed the best outcomes, but depending on the course of negative symptom subdomains, differences between the groups with regard to outcome change over time and between the subdomains. Our findings support and expand the existing evidence for the robust relationship of SA with functioning and quality of life (Messinger et al., 2011; Strauss et al., 2013; Fervaha et al., 2014; Foussias et al., 2014), and are also in line with our previous study in which higher ED was associated with global functioning in chronic patients

(Stiekema et al., 2016). In contrast to the relationship between SA and functioning, the relationship between ED and outcomes is less consistent in the literature. Our results indicate that SA is an important treatment target for improving functioning and well-being, but they also point out the importance of ED and suggest that neglecting this subgroup of negative symptoms may be disadvantageous for negative symptom treatment development.

4.3. Implications

The evidence for different relationships between the subgroups and functioning and quality of life suggests that distinguishing between the subgroups is clinically relevant and may have implications for clinical practice and treatment development. Of the included patients in this study, about half could be classified in both the low SA and low ED group. Thus, for these patients this leaves limited room (and need) for improvement in negative symptoms. This poses a problem for treatment development, as it causes an increased risk of false negative findings of treatment trials aimed at improving negative symptoms (and accompanied improvement in functioning); possibly improvement of the other 48% of the patients for whom improvement is possible and necessary is masked as the steady low group may average out effects for the other groups. This is clearly visible in our graphical representation of the overall and subgroup courses in Figures 1 and 2, where the low groups show the same pattern as the overall group but at a slightly lower level of symptoms and better level of outcomes. The finding that some of the courses of outcomes that appear visibly different were not significant (such as the differences in living situation within ED) could indicate that other factors (such as positive symptoms and neurocognition) are important for the level and course of the outcomes as well, but it could also be due to reduced power of subgroup analysis.

To date, few pharmacological studies have investigated the selective responsiveness of both subdomains, but they do not yet provide enough information for a firm conclusion about a differential response to treatment (Kirkpatrick, 2014; Azorin et al., 2014). Overall, there is only a small number of studies investigating psychosocial interventions with negative symptoms as a primary outcome (Elis et al., 2013), and the differential effects on the subdomains are unclear. SA has been related to the deficits in anticipatory pleasure (Foussias et al., 2014; Buck and Lysaker, 2013) which may make cognitive behavioral therapy a suitable intervention to address defeatists beliefs (Staring et al., 2013). ED has been linked to cognitive deficits (Liemburg et al., 2013; Bell et al., 2013; Ergul and UCok, 2015; Hartmann-Riemer et al., 2015), which may lead treatment development in the direction of restorative and/or compensatory cognitive rehabilitation interventions. However, the significant relation of ED with global functioning while controlling for cognition indicates that cognition cannot fully explain this association. Possibly, interventions targeting expressive skills such as Social Skills Training (Bellack et al., 2004; Turner et al., 2014) could improve ED. As mentioned above, studies evaluating such treatments should only include patients with more profound negative symptoms to prevent treatment effects from being masked by those with low negative symptom levels.

4.4. Strengths, limitations and future directions

Strengths of this study are the large sample size, the longitudinal nature and the used methodology. Furthermore, participants were included from representative inpatient and outpatient services covering 75% of the population in the Netherlands (Korver et al., 2012). Several limitations should also be mentioned. We cannot infer causality from this observational study. Furthermore, we do not know whether changes in negative symptoms are due to relief of secondary negative symptoms, for example by reduced positive symptoms, depressive symptoms or antipsychotic medication (Carpenter and Kirkpatrick, 2015), which may require a different approach than the suggested treatment strategies described above (Carpenter et al., 1985). Also, the current results are most applicable for those with predominant SA or ED (Strauss et al., 2013), because we have compared subgroups within each subdomain. Further, the intervals between the measures are large (3 years). Lastly, due to the relatively demanding protocol of the GROUP study, participants may differ from participants in studies that are less demanding or from patients that refuse to participate (Korver et al., 2012). Notably, in the current study non-completers showed more severe SA and ED, suggesting that patients with high levels of negative symptoms may be less likely to complete and/or participate in the study. Although this is not unique to our study, our findings should be interpreted in light of this possible selection bias.

Thus, future research should investigate possible causal mechanisms for the variability in the subdomain levels over time, e.g. whether improvement in negative symptoms facilitates improvements in outcome or vice versa (Alvarez-Jimenez et al., 2012) and whether the groups differ with regard to the care they (have) receive(d). For those with co-occurring SA and ED, research into the influence of combinations of SA and ED subgroups is needed, but this was beyond the scope of our study. Research on more specific diagnostic groups could be of value as well, since the low SA and low ED group included significantly fewer patients with schizophrenia.

4.5. Conclusion

In summary, our results show that there is a considerable heterogeneity in the course of the subdomains and suggest that negative symptoms are less stable than was previously assumed. The subgroups that we identified within SA and ED, showing a different course of symptoms over time, are clinically relevant as they are differentially related to the level and course of outcomes. Including the whole range of negative symptoms instead of distinguishing subdomains of SA and ED may explain why efforts to develop treatments for negative symptoms have been disappointing, as treatment effects may have been masked. Thus, research on treatments for negative symptoms could benefit from distinguishing subgroups within SA and ED.

References

Alvarez-Jimenez, M., Gleeson, J.F., Henry, L.P., Harrigan, S.M., Harris, M.G., Killackey, E., Bendall, S., Amminger, G.P., Yung, A.R., Herrman, H., Jackson, H.J., McGorry, P.D. (2012). Road to full recovery: longitudinal relationship between symptomatic remission and psychosocial recovery in first-episode psychosis over 7.5 years. *Psychological medicine* 42:595-606.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders*. Washington DC.

Azorin, J., Belzeaux, R., Adida, M., Hospital, S.M., Marguerite, B.S. (2014). Negative Symptoms in Schizophrenia: Where We have been and Where We are Heading. *CNS neuroscience & therapeutics* 20:801-8.

Bell, M.D., Corbera, S., Johannesen, J.K., Fiszdon, J.M., Wexler, B.E. (2013). Social cognitive impairments and negative symptoms in schizophrenia: are there subtypes with distinct functional correlates?. *Schizophrenia bulletin* 39:186-96.

Bellack, A.S., Mueser, K.T., Gingerich, S., Agresta, J. (2004). *Social skills training for schizophrenia: A step-by-step guide*. New York: Guilford Press.

Birchwood, M., Smith, J., Cochrane, R., Wetton, S., Copestake, S. (1990). The Social Functioning Scale. The development and validation of a new scale of social adjustment for use in family intervention programmes with schizophrenic patients. *The British Journal of Psychiatry* 157:653-859.

Bobes, J., Arango, C., Garcia-Garcia, M., Rejas, J. (2010). Prevalence of negative symptoms in outpatients with schizophrenia spectrum disorders treated with antipsychotics in routine clinical practice: findings from the CLAMORS study. *The Journal of clinical psychiatry* 71:280-6.

Buck, B., Lysaker, P.H. (2013). Consummatory and anticipatory anhedonia in schizophrenia: Stability, and associations with emotional distress and social function over six months. *Psychiatry research* 205:30-35.

Carpenter, W.T., Heinrichs, D.W., Alphs, L. (1985). Treatment of Negative Symptoms. *Schizophrenia bulletin* 11:440-452.

Carpenter, W.T., Kirkpatrick, B. (2015). Concepts and methods when considering negative symptom course. *Psychological medicine*1-2.

Chang, W.C., Hui, C.L.M., Tang, J.Y.M., Wong, G.H.Y., Lam, M.M.L., Chan, S.K.W., Chen, E.Y.H. (2011). Persistent negative symptoms in first-episode schizophrenia: A prospective three-year follow-up study. *Schizophrenia research* 133:22-28.

Eaton, W.W., Thara, R., Federman, B., Meton, B., Liang, K. (1995). Structure and Course of Positive and Negative Symptoms in Schizophrenia. *Archives of General Psychiatry* 52:127-134.

Elis, O., Caponigro, J.M., Kring, A.M. (2013). Psychosocial treatments for negative symptoms in schizophrenia: current practices and future directions. *Clinical psychology review* 33:914-28.

Ergul, C., UCok, A. (2015). Negative symptom subgroups have different effects on the clinical course of schizophrenia after the first episode: A 24-month follow up study. *European Psychiatry* 30:14-19.

Evensen, J., RÃ¸ssberg, J.I., Barder, H., Haahr, U., Hegelstad, W.T.V., Joa, I., Johannessen, J.O., Larsen, T.K., Melle, I., Opjordsmoen, S., Rund, B.R., Simonsen, E., Sundet, K., Vaglum, P., Friis, S., McGlashan, T. (2012). Apathy in first episode psychosis patients: A ten year longitudinal follow-up study. *Schizophrenia research* 136:19-24.

Fervaha, G., Agid, O., Takeuchi, H., Foussias, G., Remington, G. (2014). Effect of antipsychotic medication on overall life satisfaction among individuals with chronic schizophrenia: Findings from the NIMH CATIE study. *European Neuropsychopharmacology* 24:1078-1085.

Foussias, G., Mann, S., Zakzanis, K.K., van Reekum, R., Agid, O., Remington, G. (2011). Prediction of longitudinal functional outcomes in schizophrenia: the impact of baseline motivational deficits. *Schizophrenia research* 132:24-7.

Foussias, G., Siddiqui, I., Fervaha, G., Agid, O., Remington, G. (2014). Dissecting negative symptoms in schizophrenia: Opportunities for translation into new treatments. *Journal of psychopharmacology (Oxford, England)*1-11.

Galderisi, S., Bucci, P., Mucci, A., Kirkpatrick, B., Pini, S., Rossi, A., Vita, A., Maj, M. (2013). Categorical and dimensional approaches to negative symptoms of schizophrenia: focus on long-term stability and functional outcome. *Schizophrenia research* 147:157-62.

Gardner, D.M., Murphy, A.L., O'Donnell, H., Centorrino, F., Baldessarini, R.J. (2010). International consensus study of antipsychotic dosing. *The American Journal of Psychiatry* 167:686-693.

Hartmann-Riemer, M., Hager, O.M., Kirschner, M., Bischof, M., Kluge, A., Seifritz, E., Kaiser, S. (2015). The association of neurocognitive impairment with diminished expression and apathy in schizophrenia. *Schizophrenia research* 169:427-32.

Hovington, C.L., Bodnar, M., Joober, R., Malla, A., Lepage, M. (2012). Identifying persistent negative symptoms in first episode psychosis. *BMC psychiatry* 12:224-224.

Jones, B.L., Nagin, D.S., Roeder, K. (2001). A SAS Procedure Based on Mixture Models for Estimating Developmental Trajectories. *Sociological Methods & Research* 29:374-393.

Kay, S.R., Fiszbein, A., Opler, L.a. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia bulletin* 13:261-76.

Kirkpatrick, B. (2014). Developing Concepts in Negative Symptoms. *The Journal of clinical psychiatry* 75:3-7.

Korver, N., Quee, P.J., Boos, H.B.M., Simons, C.J.P., de Haan, L., GPOUPinvestigators, Investigators, G. (2012). Genetic Risk and Outcome of Psychosis (GROUP), a multi-site longitudinal cohort study focused on gene-environment interaction: objectives, sample characteristics, recruitment and assessment methods. *International Journal of Methods in Psychiatric Research* 21:205-221.

Li, K.-., Meng, X.-., Raghunathan, T.E., Rubin, D.B. (1991). Significance levels from repeated p-values with multiply-imputed data. *Statistica Sinica* 1:65-92.

Liemburg, E., Castelein, S., Stewart, R., van, d.G., Aleman, A., Knegtering, H. (2013). Two subdomains of negative symptoms in psychotic disorders: established and confirmed in two large cohorts. *Journal of psychiatric research* 47:718-25.

Messinger, J.W., Tremeau, F., Antonius, D., Mendelsohn, E., Prudent, V., Stanford, A.D., Malaspina, D. (2011). Avolition and expressive deficits capture negative symptom phenomenology: implications for DSM-5 and schizophrenia research. *Clinical psychology review* 31:161-8.

Mueser, K.T., McGurk, S.R. (2004). Schizophrenia. *Lancet* 363:2036-2072.

Niyonkuru, C., Wagner, A.K., Ozawa, H., Amin, K., Goyal, A., Fabio, A. (2013). Group-based trajectory analysis applications for prognostic biomarker model development in severe TBI: a practical example. *Journal of neurotrauma* 30:938-945.

Norman, R., Manchanda, R., Harricharan, R., Northcott, S. (2015). The course of negative symptoms over the first five years of treatment: Data from an early intervention program for psychosis. *Schizophrenia research*in press-in press.

Rocca, P., Montemagni, C., Zappia, S., PiterÃ , R., Sigaudo, M., Bogetto, F. (2014). Negative symptoms and everyday functioning in schizophrenia: a cross-sectional study in a real world-setting. *Psychiatry research* 218:284-9.

Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys*. New York: J. Wiley & Sons.

Savill, M., Banks, C., Khanom, H., Priebe, S. (2014). Do negative symptoms of schizophrenia change over time? A meta-analysis of longitudinal data. *Psychological medicine*1-15.

Staring, A.B.P., Ter Huurne, M.,B., van, d.G. (2013). Cognitive Behavioral Therapy for negative symptoms (CBT-n) in psychotic disorders: a pilot study. *Journal of Behavior Therapy and Experimental Psychiatry* 44:300-6.

Stiekema, A.P.M., Liemburg, E.J., van, d.M., Castelein, S., Stewart, R., van Weeghel, J., Aleman, A., Bruggeman, R. (2016). Confirmatory factor analysis and differential relationships of the two subdomains of negative symptoms in chronically ill psychotic patients. *PLoS ONE* 11:e0149785-e0149785.

Strauss, G.P., Horan, W., Kirkpatrick, B., Fischer, B.a., Keller, W.R., Miski, P., Buchanan, R.W., Green, M.F., Carpenter, W.T. (2013). Deconstructing negative symptoms of schizophrenia: avolition-apathy and diminished expression clusters predict clinical presentation and functional outcome. *Journal of psychiatric research* 47:783-90.

Trompenaars, F.J., Masthoff, E.D., Van Heck,G.L., Hodiamont, P.P., De Vries, J. (2005). Content validity, construct validity, and reliability of the WHOQOL-Bref in a population of Dutch adult psychiatric outpatients. *Quality of Life Research* 14:151-160.

Turner, D.T., van, d.G., Karyotaki, E., Cuijpers, P. (2014). Psychological Interventions for Psychosis :. *American Journal of Psychiatry* 171:523-538.

van Buuren, S. (2007). Multiple Imputation of Discrete and Continuous Data by Fully Conditional Specification. *Statistical methods in medical research* 16:219-242.

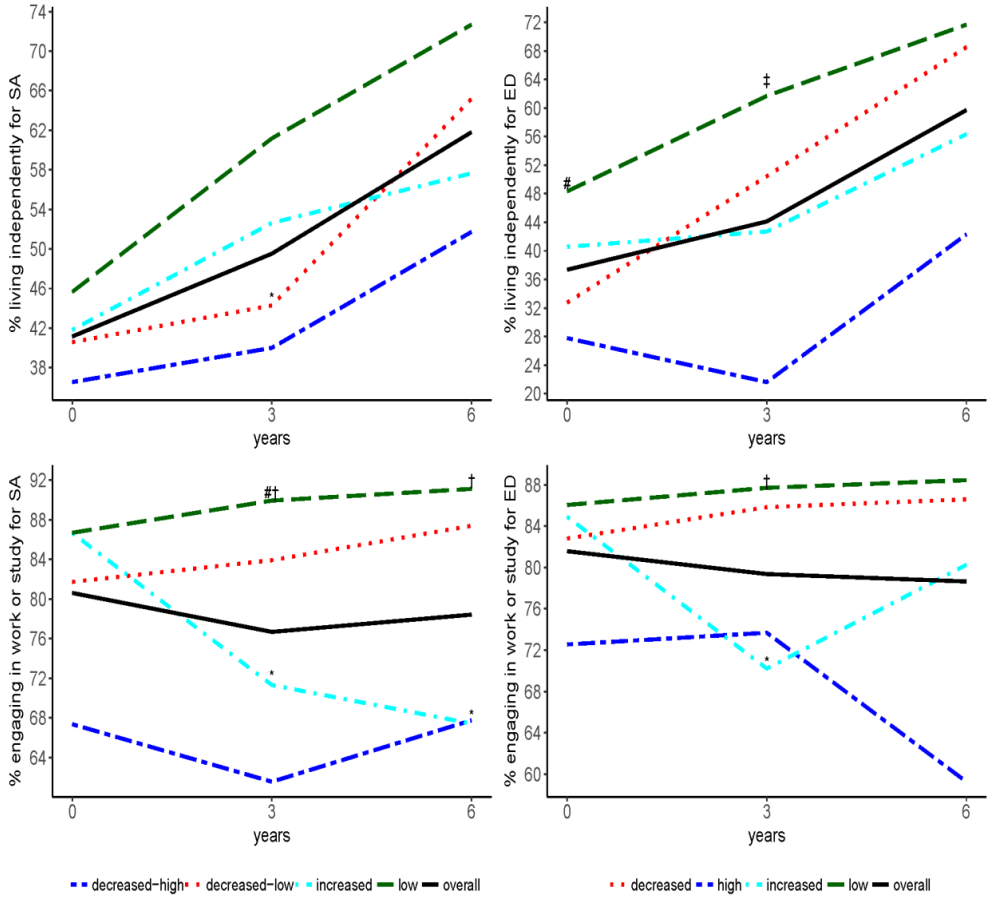
Ventura, J., Helleman, G., Thames, A.D., Koellner, V., Nuechterlein, K.H. (2009). Symptoms as mediators of the relationship between neurocognition and functional outcome in schizophrenia: a meta-analysis. *Schizophrenia research* 113:189-199.

Supplementary Materials

Supplementary table 1: Proportion of missing data for negative symptom subdomain scores and outcome variables.

	SA	ED	GAF	SFS	WHO-QOL	Work/study	Living situation
Study entry	6.2	6.7	9.1	-	11.3	10.8	7.1
3 years	29.2	29.4	32.4	31.6	31.3	27.5	28.0
6 years	43.1	43.5	45.9	45.5	45.8	40.7	44.5

Abbreviations: SA: social amotivation, ED: expressive deficits, GAF: Global Assessment of Functioning; SFS: Social Functioning Scale; WHO-QoL: World Health Organization Quality of Life.



Supplementary Figure 1: Average percentage of patients living independently and engaging in work or study (based on imputed data). Analyses were controlled for gender, duration of illness, positive symptoms and cognition. Bonferroni correction (p-value multiplied by 6) was used for pairwise comparisons. Significant differences between two groups are indicated only at the upper group. Indicated are *significantly different *course* to the low group, between study entry and the marked time point (mixed models), significantly different *level* compared to the #decreased (-low) group, †increased group and the ‡(decreased-) high group at the marked time point (pairwise comparisons).

Supplementary table 2: Bayesian Information Criterion (BIC) and logged Bayes factor (*2*ΔBIC*) for model selection*

Number of groups	BIC	*ΔBIC*	2**ΔBIC*	Evidence against H_0
Social amotivation (N=1039)				
1	-5898.15			
2	-5727.81	170.34	340.68	
3	-5654.44	73.37	146.74	
4	**-5609.96**	**44.48**	**88.96**	**Very Strong**
5	-5626.20	-16.24	-32.48	
Expressive deficits (N=1040)				
1	-6769.99			
2	-6485.22	284.77	569.54	
3	-6416.71	68.51	137.02	
4	**-6351.72**	**64.99**	**129.98**	**Very Strong**
5	-6371.48	-19.76	-39.52	

*Table 2 tabulates the BIC for model fits to the social amotivation (SA) and expressive deficits (ED) data. Based on the results the four-group model is favored for SA as well as for ED because the BIC is the smallest and 2*ΔBIC>10 suggests very strong evidence against the null model.

Supplementary table 3: Pairwise comparison using mixed models analysis adjusted for gender, duration of illness, neurocognition and positive symptoms using Bonferroni correction (p-values were multiplied by 6) and considered significant at an alpha of 0.05.

	Living situation study entry				Living situation 3 years				Living situation 6 years			
	B	SE	95% CI	p-value	B	SE	95% CI	p-value	B	SE	95% CI	p-value
Social amotivation												
decreased-high SA vs decreased-low SA	-0.53	0.64	-1.79; 0.73	1.000	-0.84	0.67	-2.15; 0.47	1.000	-0.54	0.85	-2.26; 1.17	1.000
decreased-high SA vs increased SA	-0.36	0.69	-1.71; 1.00	1.000	-1.31	0.74	-2.76; 0.14	0.464	-0.22	0.91	-2.06; 1.62	1.000
decreased-high SA vs low SA	-0.57	0.61	-1.76; 0.62	1.000	-1.61	0.67	-2.92; -0.29	0.102	-1.06	0.94	-2.99; 0.87	1.000
decreased-low SA vs increased SA	0.17	0.46	-0.74; 1.08	1.000	-0.47	0.46	-1.38; 0.44	1.000	0.33	0.53	-0.72; 1.37	1.000
decreased-low SA vs low SA	-0.04	0.33	-0.68; 0.60	1.000	-0.77	0.34	-1.43; -0.11	0.139	-0.52	0.42	-1.36; 0.32	1.000
increased SA vs low SA	-0.21	0.42	-1.04; 0.62	1.000	-0.30	0.41	-1.11; 0.52	1.000	-0.84	0.48	-1.80; 0.11	0.492
Expressive deficits												
high ED vs decreased ED	-0.23	0.65	-1.51; 1.05	1.000	-1.27	0.69	-2.64; 0.09	0.407	-1.32	0.82	-2.98; 0.34	0.698
high ED vs Increased ED	-0.63	0.68	-1.97; 0.71	1.000	-1.03	0.70	-2.41; 0.35	0.863	-0.79	0.80	-2.37; 0.80	1.000
High ED vs low ED	-1.17	0.59	-2.32; -0.03	0.271	-2.08	0.63	-3.31; -0.84	**0.006**	-1.70	0.84	-3.43; 0.02	0.317
decreased ED vs Increased ED	-0.40	0.50	-1.39; 0.58	1.000	0.24	0.51	-0.76; 1.25	1.000	0.53	0.50	-0.45; 1.52	1.000
decreased ED vs low ED	-0.94	0.35	-1.64; -0.25	**0.046**	-0.80	0.36	-1.51; -0.10	0.157	-0.38	0.42	-1.22; 0.46	1.000
Increased ED vs low ED	-0.54	0.43	-1.38; 0.30	1.000	-1.05	0.42	-1.87; -0.22	0.077	-0.92	0.46	-1.83; 0.00	0.299

	Work activities study entry				Work activities 3 years				Work activities 6 years			
	B	SE	95% CI	p-value	B	SE	95% CI	p-value	B	SE	95% CI	p-value
Social amotivation												
decreased-high SA vs decreased-low SA	-0.77	0.41	-1.58; 0.05	0.388	-0.80	0.45	-1.69; 0.10	0.488	-0.74	0.54	-1.82; 0.34	1.000
decreased-high SA vs increased SA	-0.98	0.49	-1.95; -0.02	0.278	-0.27	0.47	-1.20; 0.66	1.000	0.14	0.55	-0.96; 1.23	1.000
decreased-high SA vs low SA	-0.92	0.40	-1.71; -0.14	0.125	-1.32	0.45	-2.22; -0.42	**0.026**	-1.14	0.50	-2.13; -0.14	0.157
decreased-low SA vs increased SA	-0.22	0.40	-0.99; 0.56	1.000	0.53	0.36	-0.17; 1.23	0.838	0.88	0.34	0.22; 1.54	0.056
decreased-low SA vs low SA	-0.16	0.26	-0.66; 0.35	1.000	-0.53	0.33	-1.18; 0.13	0.677	-0.40	0.28	-0.96; 0.17	0.995
increased SA vs low SA	0.06	0.35	-0.63; 0.75	1.000	-1.05	0.32	-1.69; -0.42	**0.007**	-1.27	0.29	-1.85; -0.70	**<.001**
Expressive deficits												
high ED vs decreased ED	-0.46	0.44	-1.33; 0.40	1.000	-0.52	0.56	-1.64; 0.60	1.000	-0.86	0.51	-1.86; 0.15	0.563
high ED vs Increased ED	-0.65	0.51	-1.66; 0.35	1.000	0.14	0.57	-0.99; 1.27	1.000	-0.58	0.54	-1.66; 0.50	1.000
High ED vs low ED	-0.60	0.41	-1.42; 0.21	0.880	-0.75	0.56	-1.88; 0.38	1.000	-0.95	0.45	-1.84; -0.05	0.228
decreased ED vs Increased ED	-0.19	0.40	-0.97; 0.59	1.000	0.65	0.40	-0.13; 1.44	0.608	0.28	0.47	-0.66; 1.22	1.000
decreased ED vs low ED	-0.14	0.28	-0.69; 0.41	1.000	-0.23	0.33	-0.90; 0.43	1.000	-0.09	0.31	-0.71; 0.53	1.000
Increased ED vs low ED	0.05	0.34	-0.61; 0.71	1.000	-0.89	0.31	-1.49; -0.28	**0.025**	-0.37	0.42	-1.21; 0.48	1.000

Supplementary table 3-continued

	GAF study entry				GAF 3 years				GAF 6 years			
	B	SE	95% CI	p-value	B	SE	95% CI	p-value	B	SE	95% CI	p-value
Social amotivation												
decreased-high SA vs decreased-low SA	-4.68	2.08	-8.75; -0.61	0.146	-6.64	2.45	-11.52; -1.78	**0.047**	-5.60	2.547	-10.68; -.052	0.187
decreased-high SA vs increased SA	-6.58	2.23	-10.95; -2.21	**0.019**	0.27	2.64	-4.97; 5.50	1.000	1.26	2.667	-4.04; 6.56	1.000
decreased-high SA vs low SA	-11.92	1.97	-15.78; -8.07	**<.001**	-12.50	2.24	-16.93; -8.06	**<.001**	-9.73	2.638	-15.06; -4.40	**0.004**
decreased-low SA vs increased SA	-1.90	1.57	-4.98; 1.18	1.000	6.91	1.72	3.52; 10.29	**<.001**	6.86	1.746	3.41; -1.78	**<.001**
decreased-low SA vs low SA	-7.24	1.09	-9.38; -5.10	**<.001**	-5.85	1.19	-8.19; -3.51	**<.001**	-4.13	1.188	-6.47; -1.78	**0.004**
increased SA vs low SA	-5.34	1.38	-8.04; -2.64	**<.001**	-12.76	1.44	-15.59; -9.94	**<.001**	-10.99	1.547	-14.04; -7.93	**<.001**
Expressive deficits												
high ED vs decreased ED	-3.20	2.11	-7.34; 0.94	0.778	-5.74	2.51	-10.73; -0.74	0.150	-6.43	2.88	-12.26; -0.60	0.778
high ED vs Increased ED	-6.81	2.25	-11.23; -2.39	**0.015**	-1.20	2.49	-6.12; 3.71	1.000	-1.16	2.79	-6.74; 4.43	1.000
High ED vs low ED	-11.66	1.90	-15.38; -7.93	**<.001**	-9.67	2.36	-14.38; -4.96	**0.001**	-8.52	2.75	-14.11; -2.92	**0.024**
decreased ED vs Increased ED	-3.61	1.64	-6.82; -0.40	0.165	4.53	1.80	0.99; 8.07	0.073	5.27	1.83	1.66; 8.88	**0.026**
decreased ED vs low ED	-8.46	1.19	-10.80; -6.11	**<.001**	-3.93	1.23	-6.34; -1.53	**0.008**	-2.09	1.68	-5.52; 1.35	1.000
Increased ED vs low ED	-4.85	1.43	-7.64; -2.05	**0.004**	-8.47	1.47	-11.35; -5.58	**<.001**	-7.36	1.81	-10.98; -3.74	**0.001**

	SFS 3 years				SFS 6 years			
	B	SE	95% CI	p-value	B	SE	95% CI	p-value
Social amotivation								
decreased-high SA vs decreased-low SA	-4.83	1.53	-7.87; -1.78	**0.014**	-4.09	1.61	-7.33; -0.85	0.086
decreased-high SA vs increased SA	0.46	1.50	-2.50; 3.42	1.000	1.37	1.63	-1.88; 4.63	1.000
decreased-high SA vs low SA	-8.51	1.33	-11.14; -5.89	**<.001**	-7.36	1.48	-10.33; -4.39	**<.001**
decreased-low SA vs increased SA	5.29	1.03	3.26; 7.31	**<.001**	5.46	0.98	3.54; 7.38	**<.001**
decreased-low SA vs low SA	-3.69	0.77	-5.22; -2.16	**<.001**	-3.27	0.73	-4.71; -1.83	**<.001**
increased SA vs low SA	-8.97	0.91	-10.76; -7.19	**<.001**	-8.73	0.92	-10.54; -6.93	**<.001**
Expressive deficits								
high ED vs decreased ED	-3.68	1.52	-6.71; -0.65	0.107	-3.07	1.64	-6.36; 0.22	0.403
high ED vs Increased ED	-2.13	1.50	-5.09; 0.83	0.944	-1.50	1.63	-4.74; 1.47	1.000
High ED vs low ED	-7.05	1.40	-9.84; -4.25	**<.001**	-5.36	1.67	-8.76; -1.96	**0.018**
decreased ED vs Increased ED	1.55	1.14	-0.70; 3.80	1.000	1.57	1.28	-0.98; 4.12	1.000
decreased ED vs low ED	-3.37	0.77	-4.89; -1.84	**<.001**	-2.29	0.94	-4.19; -0.40	0.112
Increased ED vs low ED	-4.92	0.96	-6.81; -3.02	**<.001**	-3.86	1.04	-5.93; -1.79	**0.002**

Supplementary table 3-continued

	WHOQOL-BREF study entry				WHOQOL-BREF 3 years				WHOQOL-BREF 6 years			
	B	SE	95% CI	p-value	B	SE	95% CI	p-value	B	SE	95% CI	p-value
Social amotivation												
decreased-high SA vs decreased-low SA	-3.54	2.16	-7.79; 0.70	0.608	-6.44	2.34	-11.07; -1.82	**0.039**	-4.43	2.64	-9.72; 0.86	0.593
decreased-high SA vs increased SA	-3.05	2.34	-7.63; 1.54	1.000	-1.88	2.39	-6.58; 2.81	1.000	2.86	2.59	-2.25; 7.97	1.000
decreased-high SA vs low SA	-9.59	2.05	-13.61; -5.57	**<.001**	-11.92	2.12	-16.09; -7.75	**<.001**	-7.80	2.28	-12.31; -3.29	**0.005**
decreased-low SA vs increased SA	0.50	1.68	-2.80; 3.79	1.000	4.56	1.65	1.32; 7.80	**0.035**	7.29	1.89	3.54; 11.05	**0.001**
decreased-low SA vs low SA	-6.04	1.12	-8.24; -3.84	**<.001**	-5.47	1.13	-7.70; -3.25	**<.001**	-3.37	1.24	-5.82; -0.92	**0.045**
increased SA vs low SA	-6.54	1.43	-9.34; -3.74	**<.001**	-10.03	1.47	-12.92; -7.15	**<.001**	-10.66	1.53	-13.67; -7.65	**<.001**
Expressive deficits												
high ED vs decreased ED	-0.17	2.25	-4.58; 4.24	1.000	-1.88	2.32	-6.43; 2.68	1.000	-0.87	2.62	-6.10; 4.36	1.000
high ED vs Increased ED	1.13	2.33	-3.43; 5.69	1.000	-0.28	2.35	-4.90; 4.33	1.000	2.29	2.90	-3.51; 8.10	1.000
High ED vs low ED	-3.52	2.04	-7.53; 0.48	0.508	-5.28	2.09	-9.38; -1.17	0.071	-2.48	2.59	-7.70; 2.74	1.000
decreased ED vs Increased ED	1.30	1.72	-2.07; 4.67	1.000	1.59	1.82	-1.99; 5.18	1.000	3.16	1.96	-0.72; 7.03	0.655
decreased ED vs low ED	-3.35	1.21	-5.73; -0.97	**0.035**	-3.40	1.26	-5.88; -0.92	**0.044**	-1.62	1.36	-4.30; 1.07	1.000
Increased ED vs low ED	-4.65	1.44	-7.47; -1.83	**0.007**	-4.99	1.51	-7.96; -2.03	**0.006**	-4.77	1.62	-7.96; -1.58	**0.022**

Supplementary table 4: Pooled Type-III tests of fixed effects.

	Social Amotivation					Expressive Deficits				
Effect	Living situation	Work/study	GAF	SFS	WHOQOL-BREF	Living situation	Work/study	GAF	SFS	WHOQOL-BREF
	p-value	p-value	p-value	p-value	p-value	p-value	p-value	p-value	p-value	p-value
Gender	0.002	0.096	0.061	0.001	0.643	0.002	0.044	0.014	0.001	0.395
Duration of illness	<0.001	0.537	0.425	0.265	0.460	<0.001	0.609	0.215	0.113	0.318
Neurocognition	<0.001	0.001	<0.001	<0.001	0.043	<0.001	0.001	<0.001	<0.001	0.058
PANSS positive	0.008	<0.001	<0.001	<0.001	<0.001	0.019	<0.001	<0.001	<0.001	<0.001
Time	<0.001	0.066	0.001	<0.001	<0.001	<0.001	0.123	<0.001	<0.001	<0.001
SA groups	0.040	<0.001	<0.001	<0.001	<0.001	…	…	…	…	…
SA groups*time	0.053	**0.025**	**<0.001**	0.303	**0.001**	…	…	…	…	…
ED groups	…	…	…	…	…	0.001	0.005	<0.001	<0.001	<0.001
ED groups*time	…	…	…	…	…	0.170	0.236	**<0.001**	**0.030**	0.260

Part B: Statistical Analysis for Associations

CHAPTER 5

The predictive value of neurocognition and social cognition for the development of psychotic experiences in siblings of people with psychotic disorders

Md. Atiqul Islam[‡]
Gerdina H.M. Pijnenborg[‡]
Edwin R. van den Heuvel
Mark van der Gaag
Richard Bruggeman
Agna A. Bartels-Velthuis
Stefanie M. de Vries
GROUP investigators

[‡]Shared first authorship

Submitted to Journal of Neuropsychology

Abstract

Background: Neurocognitive and social cognitive impairments are associated with psychotic experiences in individuals with psychotic disorders. Diminished cognitive functioning has also been found in non-affected siblings at genetic high risk for psychosis. Our aim is to investigate the relationship of neurocognitive and social cognitive measures with course and impact of psychotic experiences in siblings of people with psychotic disorders.

Methods: Data were obtained from the Genetic Risk and Outcome of Psychosis (GROUP) project: a longitudinal multi-center cohort study in the Netherlands and Belgium. Neurocognitive and social cognitive functioning was assessed at baseline in 873 siblings (age of 35 years or less) of individuals with psychotic disorders. Frequency and distress of psychotic experiences were assessed at baseline and 3-year follow-up. A mixture of generalized linear mixed-effects model was used to test for associations.

Results: Poorer baseline verbal learning performance predicted the occurrence of psychotic experiences after three years and the distress associated with these psychotic experiences. Moreover, better baseline performance on a Theory of Mind (ToM) task was associated with a decrease of psychotic experiences over three years. Baseline distress was associated with poorer recognition of angry and neutral faces and strikingly, with better recognition of faces in general.

Conclusion: Verbal learning and ToM were found to be predictive of frequency, distress and course of psychotic experiences over three years respectively. Our findings suggest that even though cognitive functioning is poorer in people at GHR, this poorer functioning is not a robust predictor of the course of psychotic experiences.

Keywords: cognition; genetic high risk; ultra-high risk; neurocognition; social cognition; siblings; psychosis

1. Introduction

Neurocognitive and social cognitive deficits are core features of psychotic disorders and are associated with psychotic experiences (Heinrichs and Zakzanis, 1998; Doody et al., 1998; Elvevag and Goldberg, 2000), poor every day functioning and reduced quality of life (Pijnenborg et al., 2009; Green et al., 2000; Green et al., 2004; Alptekin et al., 2005; Irani et al., 2012). A milder degree of cognitive impairment may be already present in attenuated form in people at clinical high risk (CHR) for psychosis (Lencz et al., 2006; Addington et al., 2008; Becker et al., 2010; Kim et al., 2011) as well as in non-affected individuals at genetic high risk (GHR) for psychosis (Kremen et al., 1994; Appels et al., 2003; Snitz et al., 2006; Meijer et al., 2012; Quee et al., 2014). In line with this, several studies found brain abnormalities in cognitively impaired relatives at GHR (Bhojraj et al., 2011; Crossley et al., 2009), consistent with fronto-temporal dysfunctions seen in patients with schizophrenia (Crossley et al., 2009; Chua et al., 2007; Gur et al., 2000; Hirayasu et al., 2001). These findings suggest that neurocognitive and social cognitive deficits precede the onset of psychosis instead of being a consequence of psychosis (Bora et al., 2014).

The lifetime risk of developing schizophrenia in the general population is 1%, the risk of siblings at GHR thought to be approximately ten times higher (Gottesman and Wolfgram, 1991). However, certainly not all CHR or GHR individuals show an increase in psychotic experiences over time and/or make the transition to a psychotic disorder. It is suggested that cognitive functioning is associated with both the prevalence and course of psychotic experiences over time in these high risk groups.

There are several studies that examined cognitive functioning in high risk groups. Unfortunately, most studies on cognitive functioning in high risk groups are of cross-sectional design, comparing cognitive performance of CHR groups (Addington et al., 2008; Frommann et al., 2011; van Rijn et al., 2011; Myles-Worsley et al., 2007; Kim et al., 2010; Thompson et al., 2012; Chung et al., 2008) and GHR (Meijer et al., 2012; Myles-Worsley et al., 2007; Hughes et al., 2005; Klemm et al., 2006; de Achaval et al., 2010; Bertisch et al., 2008) with (first-episode) schizophrenia patients and/or healthy controls. These studies found cognitive alterations in CHR and GHR individuals, mostly at a level that is in-between healthy controls and individuals with psychotic disorders. Although these studies provide insight in the cognitive domains associated with psychosis, they do not give information on the direction of causality of the relation between cognitive functioning and increase in psychotic experiences. To this end, longitudinal studies would be needed.

However, the longitudinal studies that were done focused on the relationship between cognitive functioning and transition to psychosis as a discrete outcome and not on increased psychotic experiences as a continuous measure. By focusing exclusively on transition, information on cognitive variables that are continuously associated with an increase of psychotic symptoms is lost. Previous studies on transition showed that cognitive functioning is more impaired in those who transitioned compared with those who did not (Becker et al., 2010; Bora et al., 2014; Keefe et al., 2006; Pukrop et al., 2007; Riecher-Rossler et al., 2009; Kim et al., 2011; Agnew-Blais and Seidman, 2013), but did not consistently report specific cognitive factors to be predictive of transition to psychosis (Addington and Barbato, 2012) .

Regardless of cognitive functioning, distress caused by psychotic experiences is an important predictor of transition to psychosis and of a more severe course in general (Garety et al., 2007). However, little is known about the relationship between cognitive functioning and distress caused by psychotic experiences in at risk samples. One previous study showed hypothalamic-pituitary-adrenal (HPA) axis abnormalities, which are an indicator of distress, are associated with poorer cognitive functioning, in children with a heightened risk for psychotic disorders that, suggesting at least a cross-sectional association between distress and cognition in high risk groups (Cullen et al., 2014).

In sum, although it is assumed that certain cognitive variables may be related to vulnerability for psychosis, while others may be more related to the actual onset of the disorder (Bora et al., 2014), it is not exactly clear which cognitive variables are associated with the frequency and increase of psychotic experiences over time. In the current longitudinal cohort study these issues will be addressed in a large GHR sample.

The aim of this study is to investigate the relationship of neurocognitive and social cognitive measures with the frequency and course of psychotic experiences and distress caused by these experiences in a large GHR sample. Given that psychotic experiences are dimensional rather than dichotomous in at risk groups (van Os et al., 2009), the primary outcomes of the current study are the presence and course of psychotic experiences over time rather than transition to psychotic disorder. It is expected that poorer cognitive functioning at baseline is associated with both frequency and distress of psychotic experiences three years later and the change in psychotic experiences and associated distress over three years. Since there is hardly any literature on the association between distress caused by psychotic experiences and cognitive functioning, this association will be examined in an exploratory way.

2. Methods

2.1. Subjects

Data were collected as part of the longitudinal observational multicenter study Genetic Risk and Outcome of Psychosis (GROUP) in the Netherlands and Belgium on vulnerability and resilience in both patients with a non-affective psychotic disorder, their unaffected family members and non-related controls. The details of the GROUP study have been presented elsewhere (Korver et al., 2012). For the present study, data from baseline and 3-year follow-up were used. Siblings were asked to participate if they had at least one participating sibling with a DSM-IV (American Psychiatric Association, 2000) diagnosis of non-affective psychotic disorder. Siblings could be included if they (i) were between 16 and 50 years, (ii) had a good command of Dutch language and (iii) had no lifetime psychotic disorder at baseline, which was assessed with the Comprehensive Assessment of Symptoms and History (CASH) interview (Andreasen et al., 1992) or the Schedules for Clinical Assessment in Neuropsychiatry (SCAN 2.1) (Wing et al., 1990). For this study, all siblings with age >35 were excluded, because this generally considered the upper boarder of the age-range in which psychotic disorders usually have their onset. Siblings diagnosed with a lifetime psychotic disorder were included in the patient group. Finally, the sample at baseline (data release version 3.02) consisted of 873 unaffected siblings of individuals with psychotic disorders. For 136 families, more than one sibling of the patient was included. To examine the specificity of the descriptive results, 386

healthy control subjects were included with age ≤35 years. Controls were excluded if they had a life-time psychotic disorder of a first-degree family member with a life-time psychotic disorder. Other inclusion and exclusion criteria for controls were same as for siblings.

The GROUP study was approved by the Medical Ethics Committee of the University Medical Center Utrecht. All The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. Written informed consent was obtained from all participants.

2.2. Measures

Psychotic experiences and cognitive functioning were assessed using multiple instruments presented in a fixed order by trained research assistants (for details see (Korver et al., 2012)).

2.3. Psychotic experiences

The dependent variables were the frequency of psychotic experiences and the amount of distress of those experiences as measured by the Community Assessment of Psychic Experiences (CAPE) (Konings et al., 2006). The CAPE is a widely applied self-report questionnaire consisting of 42 items (www.cape42.homestead.com) on three dimensions: positive psychotic-like experiences (20 items), negative symptoms (14 items) and depressive symptoms (8 items) (Stefanis et al., 2002). For the present analyses, only the positive dimension was included. Subjects were asked to indicate the frequency of PE and the amount of distress resulting from these experiences on a 4-point Likert scale, ranging from 'sometimes' to 'nearly always'. Only if frequency was rated positively, the experienced distress thereof was measured on a 4-point Likert scale, ranging from 'a bit' to ‘very distressed'. A weighted mean score was calculated for psychotic experiences frequency and distress at baseline and follow-up measurement respectively. At baseline, lifetime prevalence of psychotic experiences was assessed, while at follow-up PE occurring during the last three years was assessed. Transition to psychosis between baseline and 3-year follow-up was based on a diagnosis of psychotic disorder, as assessed with the CASH or SCAN (Andreasen et al., 1992; Wing et al., 1990).

2.4. Measures of neurocognition and social cognition

The cognitive battery was based on the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) consensus (Nuechterlein et al., 2004) and included the domains of verbal learning and memory, sustained attention and vigilance, executive functions (set shifting and problem solving), visuospatial abilities, processing speed, verbal comprehension, working memory, global cognitive functioning and social cognition, including ToM and emotion perception. General face recognition was included to correct for non-emotional facial recognition (see Table 1). Baseline and 3-year follow-up measurement psychotic experiences were included, whereas for neurocognitive and social cognitive functioning only baseline measurement was analyzed.

2.5. Substance use

Baseline cannabis use was included as a confounding factor because of its association with both cognitive functioning (van der Meer et al., 2014) and PE (van Winkel and GROUP Investigators, 2015). Cannabis use (yes/no) was assessed with the Composite International Diagnostic Interview (CIDI) (WHO, 1990).

2.6. Statistical Analyses

2.6.1. Descriptive

Socio-demographic, neurocognition and social cognition characteristics of siblings at baseline were compared with healthy controls using the Mann-Whitney U test, T-test or Chi-square test depending on the type of variable.

2.6.2. Attrition and Missingness

Differences on baseline cognitive measures and PE between drop-outs and non-drop-outs for siblings and control subjects were analyzed. There were many differences between responders and non-responders on cognitive measurement. These differences are presumably all related to symptoms of psychotic disorders and therefore, we assumed that results based on complete cases would most likely be biased. The attrition rates for both outcomes were calculated as Attrition (%) = 100 − (siblings available at follow-up/siblings available at baseline based on outcomes) (Howie and Straker, 2016). To reduce attrition with this young group of siblings, we dealt with missing data in statistical way where attrition is difficult to avoid (Hansen et al., 1985; Davis et al., 2002; Horton and Kleinman, 2007; Howie and Straker, 2016). The missing values in the outcomes and independent variables arose from a combination of absenteeism, attrition, or a failure to complete the questionnaire on time. In this study, the missing data patterns showed approximately one-third of the individuals responded on all dependent and independent variables at baseline and three years. Ignoring missing data yielded biases as it does not differentiate missing at random mechanism (Little and Rubin, 2002). Therefore, multiple imputation was applied to address missing data in outcomes and independent variables. A fully conditional specification (FCS) predicted mean matching (PMM) method was used to impute missing values for both continuous and categorical variables (van Buuren, 2007), and 20 imputed datasets each with sample size of 873 siblings were generated. All variables, including socio-demographic (age, gender, ethnicity and education), cannabis use, neurocognitive and social cognitive measures as well as the dependent variables at both time points were included in the imputation model. Subsequently, each dataset was analyzed with the appropriate statistical model (e.g. mixed effects models). Parameter estimates and their associated variances were pooled with Rubin's rule (Rubin, 1987).

2.6.3. Mixture distribution model

Observations within siblings were not independent (repeated measures). Therefore, generalized linear mixed-effects model was used to estimate the unique association between measures of neurocognition and social cognition and the development of psychotic experiences. Since our two outcome variables included only zeros (many siblings did not report symptoms) or non-negative

values, these distribution of these variables are extremely skewed (Supplementary Figure S1-S2). Therefore, the outcome variables (continuous part) were modeled with a lognormal distribution allowing random effects. Subsequently, a logistic regression with random effect model was used for estimating Bernoulli probability of a true nonzero score on the frequency of psychotic experiences or the amount of distress of those psychotic experiences in siblings. More specifically, intensity part of the outcome modeled with generalized linear mixed effects model (i.e. random effects lognormal model) and binary parts were modeled with generalized linear mixed effects model. These two models were to be analyzed at the same time. Therefore, we combined these two forming it as mixture distribution model or mixture of generalized linear mixed effects model (Tooze et al., 2002), allowing random intercepts (e.g. siblings) in both binary and continuous parts. We considered both correlated and uncorrelated mixture of generalized linear mixed-effects model to analyze the data. The mathematical expression and the estimation process are summarized in the method section of the supplementary materials. Briefly, the correlated mixed-distribution model was maximized by using quasi-Newton optimization of a likelihood approximated by adaptive Gaussian quadrature (Molenberghs and Verbeke, 2006; Tooze et al., 2002; Zeger and Karim, 1991). Two parts of the uncorrelated mixed distribution models were maximized separately by adaptive Gaussian quadrature. A SAS macro MIXCORR (Tooze et al., 2002) was used to fit the correlated and uncorrelated model. Within the MIXCORR macro, we fitted the models using PROC GENMOD and PROC NLMIXED and the results were used as starting values for the final estimation of the model parameters for both correlated and uncorrelated mixed effects models using PROC NLMIXED (Tooze et al., 2002).

The selected independent variables (Table 1) in the statistical model including time (categorical) were all considered as fixed effects. Since we were interested in change in PE frequency and distress, the interaction of the cognitive variables with time was our main focus. Generalized linear mixed model and linear mixed model were used to select the candidate independent variables for binary and continuous parts respectively using pooled type-III analysis of fixed effects p-values.

Two sets of variables were identified in the candidate independent variables selection model. The first set contained the confounding factors (age, gender, education, and cannabis use), which were always included in every model selection step. The second set contained the neurocognitive and social cognitive measures, which were eliminated one by one. The process continued with a backward elimination procedure on the pooled type-III analysis of fixed effects p-values until all remaining independent variables were significant at α=0.05. The main effects cannot be excluded if they interact with time. Finally all the selected candidate independent variables were used in the proposed mixture of generalized linear mixed effects model (correlated and uncorrelated) to estimate the model parameters. All analyses were done using two-tailed tests at a 5% significance level. The statistical analyses were performed using Statistical Analysis System, version 9.4.

Table 1: Measures of neuro and social cognition

Cognitive domain	Test	Outcome measure	Reference
Neurocognition			
Verbal learning and memory	Word Learning Task (WLT)	Immediate recall (total score of three15-word learning trials) and retention rate (delayed recall /maximum score immediate recall)	(Brand and Jolles, 1985)
Sustained attention and vigilance	Continuous Performance Test (CPT-HQ)	Sensitivity Index (number of correct detections of targets minus the number of false alarms for non-target stimuli	(Nuechterlein and Dawson, 1984)
Set Shifting	Response Shifting Task (RST)	Accuracy Cost Score (proportion correct in the imitation condition minus proportion correct in the reversal condition)	(Bilder et al., 1992; Meiran et al., 2000)
Global cognitive functioning	Wechsler Adult Intelligence Scale (WAIS-III, short form)	IQ	(Wechsler, 1997)
Problem solving and visuospatial abilities	Block Design	Total raw score (0-68)	
Speed of processing	Digit Symbol Coding	Total raw score (0-133)	
Verbal comprehension	Information	Total raw score (0-28)	
Working memory	Arithmetic	Total raw score (0-22)	
General face recognition	Benton Facial Recognition Test (BFRT)	Total amount of correctly matched faces	(Benton et al., 1983)
Social cognition			
Theory of Mind	Hinting Task	Total score (0-20)	(Versmissen et al., 2008)
Emotion perception	Degraded Facial Affect Recognition Task (DFAR)	The proportion of correctly identified happy, fearful, angry and neutral faces and the proportion of total correctly identified faces	(van 't Wout et al., 2004)

3. Results

3.1. Sample

At baseline, the total sample consisted of 873 siblings and 386 controls. At baseline, 757 (86.7%) siblings and 364 (94.3%) controls completed the CAPE frequency scale and 651 (74.6%) siblings and 278 (72.02%) controls at 3-year follow-up. The attrition rates for siblings and controls on PE frequency were 14% and 23.6% respectively. For the CAPE distress scale, numbers were 748 (85.7%) siblings and 364 (94.3%) controls at baseline and 649 (74.3%) siblings and 276 (71.5%) controls at 3-year follow-up. Similarly, the attrition rates for siblings and controls on PE distress were 13.2% and 24.2% respectively. In total, seven siblings (0.8%) and two controls (0.5%) made the transition to psychosis between baseline and follow-up. It should be considered that about 11.6% at baseline and 1.1% at 3-year follow-up of the siblings of the present study had mood disorders. They were all included as siblings or controls in the analyses. At baseline, none of the siblings and controls were treated with antipsychotics or other psychotropic medication (e.g. antidepressants), while at 3-year follow-up six of the siblings who transitioned to psychosis used antipsychotics. None of the controls used anti-depressive or anti-psychotic medication. The mean duration between baseline cognitive testing and follow-up for siblings and controls were 39.4 months (SD=5.6) and 39.2 months (SD=5.5) respectively.

3.2. Descriptive analysis

Table 2a demonstrates the demographic characteristics and neurocognitive and social cognitive performance at baseline and their comparison between siblings and controls. Siblings' mean age was significantly higher, their mean scores on tests of verbal learning, IQ, problem solving and visuospatial abilities, speed of processing, verbal comprehension, working memory, ToM and emotion perception (DFAR Percent of angry face) were significantly lower than that of controls at baseline.

Sibling who dropped-out showed poorer cognitive performance at baseline on verbal learning and memory ($p<0.05$), problem solving and visuospatial skills ($p<0.001$), speed of processing ($p=0.001$), verbal comprehension ($p<0.001$), working memory ($p<0.001$) and had a lower IQ ($p<0.001$) (see Supplementary Table S1-S2 for details). Drop-outs reported less or equal psychotic experiences at baseline compared with non-drop-outs, but this effect was not significant (Supplementary Table S2). In controls, drop-outs reported more psychotic experiences at baseline than non-dropouts, and the effects on psychotic experiences frequency was significant ($p=0.011$) but not significant on psychotic experiences distress (Supplementary Table S1). The relationships (Spearman correlation coefficients) between all neuro- and social cognition measurements and their significance level for both controls and siblings were presented in the Supplementary Table S3-S4.

A large majority of both siblings (87.9%) and controls (85.4%) reported psychotic experiences at baseline (ranging from 'sometimes' to 'nearly always'), of whom more than half of both siblings (57.9%) and controls (58.2%) indicated that the psychotic experiences resulted in emotional distress (ranging from 'a bit' to 'very distressed') respectively (see Table 2b). The percentage of both siblings and controls with psychotic experiences was significantly lower at follow-up than at baseline, as was the percentages of people experiences distress from these symptoms (see Table 2b).

Table 2a: Demographic, neuro and social cognition characteristics at baseline[‡]

Variable/group	Controls (N=386)		Siblings (N=873)	
	Mean±SD or (%)	Missing N (%)	Mean±SD or (%)	Missing n (%)
Age in years, mean±SD	23.81±5.50	0	24.92±5.33***	0
Gender, Female %	52.33	0	54.30	0
Ethnicity, Caucasian %	89.81	13 (3.37)	82.53	3 (0.34)
Education[a] , mean±SD	5.16±1.78	1 (0.26)	4.95±2.09	16 (1.83)
Cannabis use past 12 months, Yes %	20.68	4 (1.04)	21.25	7 (0.80)
Neurocognition, mean±SD:				
Immediate Recall[b]	29.15±5.04	6 (1.55)	27.19±5.62***	25 (2.86)
Retention Rate[c]	0.83±0.15	12 (3.11)	0.84±0.17	36 (4.12)
Sensitivity Index[d]	97.80±6.85	35 (9.07)	96.72±10.39	85 (9.74)
Accuracy Cost Score[e]	0.12±0.19	37 (9.59)	0.14±0.22	86 (9.85)
IQ[f]	109.36±14.52	6 (1.55)	102.66±15.54***	33 (3.78)
Block Design[g]	48.66±13.52	5 (1.30)	45.50±14.89***	21 (2.41)
Digit Symbol Coding[h]	85.12±13.98	5 (1.30)	80.13±15.31***	18 (2.06)
Information[i]	18.55±4.77	5 (1.30)	16.73±5.23***	19 (2.18)
Arithmetic[j]	15.15±4.21	5 (1.30)	13.79±4.46***	19 (2.18)
Benton Facial[k]	23.24±2.01	8 (2.07)	23.21±2.16	26 (2.98)
Social Cognition, mean±SD:				
Hinting Task	19.10±1.21	10 (2.59)	18.80±1.70**	24 (2.75)
Degraded Facial Affect Recognition (DFAR):				
Percent happy faces	88.83±10.69	30 (7.77)	88.40±10.65	66 (7.56)
Percent fearful faces	56.02±17.50	30 (7.77)	54.52±19.19	66 (7.56)
Percent angry faces	71.52±18.81	30 (7.77)	69.42±18.98*	66 (7.56)
Percent neutral faces	81.69±14.87	30 (7.77)	80.23±15.15	66 (7.56)

[‡]Table presents mean±SD or number (%); [a]Education (Verhage, 1965): range 0 (no education), 3-5 (school diploma), 6-8 (professional education/university degree); [b]Immediate Recall: WLT Immediate Recall; [c]Retention rate: WLT Retention Rate; [d]Sensitivity Index: Continuous Performance Test Sensitivity Index; [e]Accuracy Cost Score: Response Shifting Task Accuracy Cost Score; [f]IQ: WAIS-III Intelligence Quotient; [g]Block Design: WAIS-III Block Design; [h]Digit Symbol Coding: WAIS-III Digit Symbol Coding; [i]Information: WAIS-III Information; [j]Arithmetic: WAIS-III Arithmetic; [k]Benton Facial: Benton Facial Recognition Test. Significance levels: ***p <0 .001, **p <0 .01 and *p < 0.05.

Table 2b: Descriptive statistics for the outcomes at baseline and three years follow-up[‡]

Outcomes	Controls (n=386)				Siblings (n=873)			
	Baseline		Follow-up		Baseline		Follow-up	
	Mean±SD	Missing %	Mean±SD	Missing %	Mean±SD	Missing %	Mean±SD	Missing %
PE frequency	0.21±0.19	5.7	0.10±0.13	27.98	0.22±0.21	13.29	0.12±0.19	25.43
PE distress	0.40±0.44	5.7	0.30±0.46	28.50	0.41±0.47	14.32	0.39±0.57	25.66
People with/without symptoms								
	With	Without	With	Without	With	Without	With	Without
	Mean±SD	%	Mean±SD	%	Mean±SD	%	Mean±SD	%
PE frequency	0.24±0.18	14.56	0.15±0.14	36.69	0.25±0.20	12.15	0.18±0.20	32.72
PE distress	0.68±0.38	41.76	0.77±0.42	61.23	0.71±0.42	42.11	0.91±0.54	57.63

[‡]Outcome variables were measured by the Community Assessment of Psychic Experiences (CAPE); table presents mean±SD and percentage of missingness and without symptoms at each time point.

3.3. Neurocognitive and social cognitive predictors for symptom development

For both outcomes with and without correlated random effects models was fitted using MIXCORR macro. For the score of the frequency of psychotic experiences, only four datasets out of 20 passed the convergence, so we pooled the parameter estimates based on the existing imputed datasets (04) which were estimable with MIXCORR macro (Supplementary Table S5). Due to the problem of convergence and correlation coefficient between the random effects of occurrence and intensity of psychotic experiences frequency that was very high (ρ=0.941), a single random effect instead of two was the best way to analyze the data. So, we fitted the two-part model incorporating a single random intercept for the intensity part and a scale parameter. The random effect of binary part was then calculated using the random effect of intensity part and the scale parameters for binary part.

Table 3 shows the pooled parameter estimates derived from the mixture of generalized linear mixed effects model for the frequency of psychotic experiences. The psychotic experiences frequency scores indicated a significant interaction effect between the score of Immediate Recall and time at three years in the occurrence part (p=0.024) and the score of Hinting Task and time at three years in the intensity part (p=0.020) respectively. A higher score on Immediate Recall at baseline was associated with an average decrease (OR=*exp*(-0.068)=0.93) of having psychotic experiences in siblings three years later. A higher score for Hinting Task at baseline was associated with an average decrease (-0.053) in psychotic experiences frequency three years later. A significant main effect was found for the score on the Degraded Facial Affect Recognition Task (DFAR) percentage of neutral faces for the intensity of the frequency of psychotic experiences. A higher score for the DFAR recognition of neutral faces was associated with less psychotic experiences at baseline. DFAR recognition score did not have an effect on a change in psychotic experiences frequency over time (absence of interaction effect). A lower score on the sensitivity index of the Continuous Performance Test-HQ and a higher score for Benton Facial Recognition Test (were associated with more psychotic experiences frequency at trend level, not reaching significance). The variance of common random intercept was 0.325, which indicated siblings observed 32.5% of the variation on both the occurrence and the intensity of score of psychotic experiences frequency.

Table 4 shows the pooled parameter estimates derived from the mixture of generalized linear mixed effects model for the amount of distress of those with psychotic experiences. The model was fitted with and without correlated random effects and correlated random effect model was better than the uncorrelated random effect model based on pooled AIC (3127.97) and log likelihood ratio test (p=0.004). The interaction effect between the score on Immediate Recall and time at three years was significant for occurrence of PE (p=0.027). A higher score of Immediate Recall (indicating better verbal learning) at baseline was associated with a decrease (OR=*exp*(-0.056)=0.94) of experiencing distress by siblings with psychotic experiences three years later. DFAR percentages of angry faces, neutral faces and BFRT showed significant main effects for the occurrence of distress of psychotic experiences but these cognitive variables were consistent over time. Percentages of angry faces and neutral faces showed negative association, and BFRT showed positive association of having psychotic experiences distress. There were no significant additional main effects for the intensity of the experienced distress of psychotic experiences, although a trend (P=0.08) was seen of the interaction of neutral faces and time at three years for intensity of psychotic experiences distress. The random

effects in the correlated mixed-distribution model $\sigma_1^2$and σ_2^2 explained the unobserved heterogeneity among siblings (Tooze et al., 2002). The variance of the random effect $\sigma_1^2 = 2.57$ indicated the high variability of the probability of a nonzero psychotic experiences distress among siblings with similar covariate patterns, and $\sigma_2^2 = 0.08$ allowed to have consistently low mean of nonzero values of psychotic experiences distress. The correlation coefficient between the random effects of occurrence and intensity of the experienced distress of psychotic experiences was 43.8%, indicating consistently high occurrence probability to have consistently high mean of nonzero values of psychotic experiences (Table 4).

Table 3: Pooled parameter estimates and fit statistics for the frequency of positive psychotic experiences*

Parameter	Estimate	S.E.	95% C.I		P-value
			Lower	Lower	
Occurrence (Logistic)					
Intercept	1.849	1.314	-0.731	4.429	0.1598
Age	-0.009	0.023	-0.054	0.037	0.7027
Sex	0.083	0.213	-0.335	0.500	0.6982
Education	-0.078	0.058	-0.192	0.036	0.1818
Cannabis Use	-0.650	0.274	-1.188	-0.112	0.0179
Benton facial[a]	0.068	0.041	-0.013	0.148	0.0982
Immediate Recall[b]	0.041	0.026	-0.009	0.091	0.1076
Time (3 years)	0.061	0.841	-1.594	1.716	0.9423
Immediate Recall*Time (3 years)	-0.068	0.030	-0.127	-0.009	0.0240
Var(Random Effect): $\eta^2\sigma_1^2$	2.997	0.669	1.683	4.311	<.0001
Scale Parameter (η)	3.038	0.367	2.317	3.760	<.0001
Intensity (Lognormal)					
Intercept	-0.889	0.463	-1.801	0.022	0.0558
Age	-0.002	0.006	-0.015	0.010	0.7066
Sex	0.028	0.056	-0.082	0.139	0.6168
Education	-0.042	0.016	-0.073	-0.011	0.0084
Cannabis Use	-0.137	0.068	-0.270	-0.004	0.0440
Sensitivity Index[c]	-0.005	0.003	-0.011	0.000	0.0649
Percent neutral faces[d]	-0.004	0.002	-0.007	0.000	0.0307
Hinting Task	0.019	0.018	-0.016	0.054	0.2813
Time (3 years)	0.505	0.432	-0.347	1.358	0.2438
Hinting Task*Time (3 years)	-0.053	0.023	-0.099	-0.008	0.0204
Var(Random Effect): σ_1^2	0.325	0.036	0.254	0.396	<.0001
Var (Residual): σ_e^2	0.115	0.018	0.080	0.149	<.0001
Fit Statistics					
Criterion (Pooled)	Value				
AIC	-370.383				
-2LL	-414.383				

*The model adjusted for age, gender, education, and cannabis used in past 12 months; [a]Benton Facial: Benton Facial Recognition Test; [b]Immediate Recall: WLT Immediate Recall; [c]Sensitivity Index: Continuous Performance Test Sensitivity Index; [d]Percent neutral faces: Degraded Facial Affect Recognition Task percent neutral faces; AIC: Akaike Information Criterion; -2LL: -2Log Likelihood.

Table 4: Pooled parameter estimates and model comparison for the amount of experienced distress of positive psychotic experiences*

Parameter	Uncorrelated Model					Correlated Model				
	Estimate	S.E.	95% C.I		P-value	Estimate	S.E.	95% C.I		P-value
			Lower	Upper				Lower	Upper	
Occurrence (Logistic)										
Intercept	0.074	1.255	-2.391	2.538	0.9532	0.074	1.253	-2.386	2.535	0.9527
Age	-0.031	0.020	-0.071	0.009	0.1335	-0.031	0.020	-0.070	0.009	0.1336
Sex	0.792	0.197	0.404	1.181	<.0001	0.791	0.198	0.402	1.179	<.0001
Education	-0.115	0.053	-0.219	-0.010	0.0314	-0.117	0.053	-0.222	-0.013	0.0279
Cannabis Use	0.084	0.235	-0.377	0.545	0.7204	0.085	0.234	-0.375	0.546	0.7153
Percent angry faces[a]	-0.010	0.005	-0.019	0.000	0.047	-0.010	0.005	-0.020	-0.001	0.0377
Percent neutral faces[b]	-0.016	0.006	-0.027	-0.004	0.0064	-0.016	0.006	-0.027	-0.005	0.0055
Benton facial[c]	0.089	0.044	0.003	0.175	0.0414	0.088	0.044	0.001	0.174	0.0462
Time (3 years)	0.537	0.693	-0.829	1.902	0.4397	0.539	0.692	-0.824	1.903	0.4366
Immediate Recall[d]	0.008	0.020	-0.030	0.047	0.6776	0.012	0.020	-0.027	0.050	0.5531
Immediate Recall*Time (3 years)	-0.056	0.025	-0.105	-0.006	0.0268	-0.055	0.025	-0.105	-0.006	0.0271
Var(random intercept): σ_1^2	2.555	0.538	1.496	3.613	<.0001	2.566	0.555	1.474	3.658	<.0001
Intensity (Lognormal)										
Intercept	-0.722	0.220	-1.154	-0.290	0.0011	-0.726	0.222	-1.161	-0.291	0.0011
Age	-0.001	0.005	-0.011	0.010	0.9836	-0.001	0.005	-0.011	0.010	0.8705
Sex	0.118	0.050	0.020	0.216	0.0179	0.132	0.052	0.031	0.233	0.0109
Education	-0.015	0.014	-0.042	0.011	0.2572	-0.019	0.014	-0.046	0.009	0.178
Cannabis Use	0.006	0.062	-0.115	0.127	0.9243	0.003	0.062	-0.118	0.125	0.9551
Time (3 years)	0.653	0.248	0.163	1.143	0.0092	0.650	0.248	0.161	1.138	0.0095
Percent neutral faces	0.001	0.002	-0.003	0.005	0.6464	0.000	0.002	-0.003	0.004	0.8109
Percent neutral faces*Time (3 years)	-0.005	0.003	-0.011	0.001	0.0922	-0.005	0.003	-0.012	0.001	0.0808
Var(res): σ_e^2	0.311	0.034	0.244	0.379	<.0001	0.311	0.034	0.243	0.379	<.0001
Var(random intercept): σ_2^2	0.070	0.028	0.014	0.126	0.015	0.076	0.030	0.017	0.135	0.0123
Covariance	...	...	...	...	...	0.193	0.101	-0.006	0.392	0.0571
Correlation (ρ)	...	...	...	...	...	0.438				
Model comparison (Fit Statistics)										
Criterion (Pooled)	Value					Value		Difference in -2LL		P-value
AIC	3134.30					3127.97				
-2LL	3090.30					3081.97		8.33		0.004

*The model adjusted by age, gender, education, and cannabis used in past 12 months; [a]Percent angry faces: Degraded Facial Affect Recognition Task percent angry faces. [b]Percent neutral faces: Degraded Facial Affect Recognition Task percent neutral faces; [c]Benton Facial: Benton Facial Recognition Test; [d]Immediate Recall: WLT Immediate Recall; AIC: Akaike Information Criterion; -2LL: -2Log Likelihood.

4. Discussion

In the current study, we examined the predictive value of neurocognition and social cognition for the development of psychotic experiences over time in a large sample of siblings at genetic high risk (GHR) for psychotic disorder. At baseline 87.9% of the siblings reported psychotic experiences, of which 57.9% indicated that the psychotic experiences resulted in distress. At follow-up 67.3% of the siblings reported psychotic experiences in the past three years, of which 42.4% stated feelings of distress. Interestingly, neither frequency nor distress of psychotic experiences in our GHR sample differed from healthy controls. Moreover, the percentage of people that made a transition to psychosis over three years was relatively low in both groups (seven siblings (0.8%) and two controls (0.5%)). We found a substantial decrease in the frequency and distress of psychotic experiences between baseline and 3-year follow-up in both groups. This decrease may be due to the fact that baseline assessment of frequency psychotic experiences was lifetime frequency and follow-up assessment was over the three years before assessment. The high number of controls reporting at least one psychotic experience at some point in life, which is in line with previous findings in a large cohort of Dutch early adolescents in the general population, reporting a prevalence of 95% of psychotic experiences (Wigman et al., 2011). Participants in that study were younger than those in our sample (age 12-16), which may account for the fact that prevalence of psychotic experiences in this study is even somewhat higher than in our samples.

Siblings mean age was a little higher than controls (1.2 years), they had a lower mean IQ than controls and performed worse on tests of several other cognitive domains: verbal learning, visuospatial abilities, processing speed, verbal comprehension, recognition of facial affect, ToM and working memory. The widespread pattern of poorer cognitive functioning implicates that the cognitive profile of the current GHR sample is comparable to that of subjects at UHR who also show impairments in multiple cognitive domains (Fusar-Poli et al., 2012).

While baseline performance on tasks assessing verbal learning was not associated with baseline psychotic experiences frequency, poorer performance on this task predicted the occurrence psychotic experiences after three years and the distress associated with these psychotic experiences. Moreover, better baseline performance on a ToM task was associated with a decrease of psychotic symptoms after three years. Baseline distress was associated with poorer recognition of angry and neutral faces and, strikingly, with better recognition of faces in general.

Thus, in particular verbal learning and social cognitive measures were associated with frequency and distress of psychotic experiences in this GHR sample. In subjects at UHR, a broad range of cognitive functions was associated with conversion (Fusar-Poli et al., 2012; van Donkersgoed et al., 2015), among which also verbal memory and ToM. Verbal memory was among the most impaired cognitive domains in CHR and more impaired in converters than in non-converters (Seidman et al., 2016). Poorer verbal memory may make people more vulnerable for psychotic experiences because of misinterpretations of past situations and biases in the encoding of new information. Previous studies found verbal memory to be a robust predictor of social functioning as well (Green, 1996). It may be people with poorer verbal memory are less able to generate social support in case of psychotic experiences and therefore experience more distress.

ToM was the only social cognitive domain found to be associated with conversion in a previous meta-analysis (van Donkersgoed et al., 2015). In line with this, a study on the course of psychotic symptoms over time in children showed that poorer ToM predicted delusions after three year in children with auditory hallucinations (Bartels-Velthuis et al., 2011). Apparently the ability to infer mental state of others protects against psychotic experiences, probably since information on the mental state of others can be used to correct one's own thoughts and beliefs (Pijnenborg et al., 2013).

The present study has a number of limitations. First, the relatively long duration of GROUP (three years for the present study) might have the disadvantage (like other longitudinal studies on psychiatric disorders) that only more stable or better-functioning siblings will continue participating over time, while siblings who increase in psychotic experiences over time or even convert to psychosis, may have been more likely to drop out. Although there are many possible reasons for drop-out and the use of multiple imputation should have corrected substantially for its effect. It cannot be ruled out that these drop-outs may have developed more severe symptoms or may have converted to a psychotic disorder between baseline and 3-year follow-up, which could have been a reason for not being able/willing to participate. Therefore, the assessment of psychotic experiences at 3-year follow-up might have been biased, resulting in the observed decrease in psychotic experiences frequency and distress. Possibly, these drop-outs (non-responders) are the siblings of special interest, because they may have made the transition to psychosis. Indeed, it was shown that drop-outs at 3-year follow-up reported non-significantly more psychotic experiences at baseline compared with participants at 3-year follow-up. Furthermore, drop-outs at follow-up were already more neurocognitively affected at baseline compared with responders. That is, non-responders at follow-up showed decreased performance at baseline on verbal learning, IQ, problem solving and visuospatial skills, speed of processing, acquired knowledge and working memory. Second, it should be considered that about 11.6% of the siblings of the present study had mood disorders at baseline, which may have affected cognitive functioning (Baune et al., 2010). However, the fact that both mood disorders and cognitive impairments were more prevalent in siblings, also illustrates that GHR is associated with several additional risk factors that may be a cumulative risk for poor outcome in terms of symptoms. We decided not to exclude siblings suffering from mood disorders, since depression is frequently observed in schizophrenia (Buckley et al., 2009) and is found to be a predictor for psychosis (Yung et al., 2003). Moreover, a sample including siblings with mood disorders is more representative for this high risk population. Finally, psychotic experiences were assessed with a self-report questionnaire. This may have biased the results as not all psychotic experiences may be labeled as such by the participants, given that these are very personal experiences and thoughts. Reporting these experiences requires a certain level of self-reflection.

A strength of this study is the large sample size and the use of a neurocognitive test battery in which most of the cognitive domains suggested by the MATRICS consensus are incorporated (Nuechterlein et al., 2004), yielding a broad range of neurocognitive variables. Moreover, this study combines both neurocognition and social cognition, providing a comprehensive picture of the possible factors related to psychosis. Another strength is that we investigated the association between cognition and the development of psychotic experiences specifically over time, contrary to

most previous studies comparing performance of high risk subjects with healthy controls and/or individuals with psychotic disorders (Addington et al., 2008; Kim et al., 2010; Thompson et al., 2012; Laurent et al., 2001). In addition, the current study focused at baseline on non-help seeking, siblings of individuals with a psychotic disorder. This has the advantage, compared with studies on CHR groups, that cognitive performance was not confounded by the effects of antipsychotic medication or illness (duration) at the time of assessment.

In conclusion, the current study highlighted the importance of a further differentiation within a genetic high risk group based on neuro- and social-cognition. We found lower scores on tests of IQ verbal learning, visuospatial abilities, processing speed, verbal comprehension, recognition of facial affect, ToM and working memory may represent general vulnerability factors for psychotic experiences associated with their genetic liability, and may therefore not be specifically predictive of the development of psychosis. However, only a few cognitive domains were associated with frequency and distress of psychotic experiences at baseline in this study. In addition, cognitive domains that were associated with PE at baseline do not necessarily predict occurrence after three years, nor increase in psychotic experiences frequency and distress over time in people at genetic high risk for psychoses. In general, the fact that psychotic experiences were highly prevalent in both people at GHR and healthy controls may explain the lack of associations. Given that psychotic experiences are apparently widespread under healthy young adults, these may not be associated with impaired functioning in other domains, such as cognitive functioning. Only verbal learning and ToM were found to be predictive of respectively frequency and, distress and change in frequency of psychotic experiences after three years. Our findings suggest that even though cognitive functioning is poorer in people at GHR, this poorer functioning is not a robust predictor of the course of psychotic experiences.

References

Addington, J., Barbato, M. (2012). The role of cognitive functioning in the outcome of those at clinical high risk for developing psychosis. *Epidemiology and psychiatric sciences* 21:335-342.

Addington, J., Penn, D., Woods, S.W., Addington, D., Perkins, D.O. (2008). Facial affect recognition in individuals at clinical high risk for psychosis. *The British journal of psychiatry : the journal of mental science* 192:67-68.

Agnew-Blais, J., Seidman, L.J. (2013). Neurocognition in youth and young adults under age 30 at familial risk for schizophrenia: a quantitative and qualitative review. *Cognitive neuropsychiatry* 18:44-82.

Alptekin, K., Akvardar, Y., Kivircik Akdede, B.B., Dumlu, K., Isik, D., Pirincci, F., Yahssin, S., Kitis, A. (2005). Is quality of life associated with cognitive impairment in schizophrenia?. *Progress in neuro-psychopharmacology & biological psychiatry* 29:239-244.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders*. Washington DC.

Andreasen, N.C., Flaum, M., Arndt, S. (1992). The Comprehensive Assessment of Symptoms and History (CASH). An instrument for assessing diagnosis and psychopathology. *Archives of General Psychiatry* 49:615-623.

Appels, M.C., Sitskoorn, M.M., Westers, P., Lems, E., Kahn, R.S. (2003). Cognitive dysfunctions in parents of schizophrenic patients parallel the deficits found in patients. *Schizophrenia research* 63:285-293.

Bartels-Velthuis, A.A., Blijd-Hoogewys, E.M., van Os, J. (2011). Better theory-of-mind skills in children hearing voices mitigate the risk of secondary delusion formation. *Acta Psychiatrica Scandinavica* 124:193-197.

Baune, B.T., Miller, R., McAfoose, J., Johnson, M., Quirk, F., Mitchell, D. (2010). The role of cognitive impairment in general functioning in major depression. *Psychiatry research* 176:183-189.

Becker, H.E., Nieman, D.H., Wiltink, S., Dingemans, P.M., van de Fliert, J.R., Velthorst, E., de Haan, L., van Amelsvoort, T.A., Linszen, D.H. (2010). Neurocognitive functioning before and after the first psychotic episode: does psychosis result in cognitive deterioration?. *Psychological medicine* 40:1599-1606.

Benton, A.L., Sivan, A.B., Hamsher, K., Varney, N.R., Spreen, O. (1983). *Benton test of facial recognition.* New York: Oxford: University Press.

Bertisch, H.C., Fava, J., Kattan, A., DeLisi, L.E. (2008). Preliminary neuropsychological findings in individuals at high genetic risk for schizophrenia. *Early intervention in psychiatry* 2:45-49.

Bhojraj, T.S., Francis, A.N., Montrose, D.M., Keshavan, M.S. (2011). Grey matter and cognitive deficits in young relatives of schizophrenia patients. *NeuroImage* 54 Suppl 1:S287-92.

Bilder, R.M., Turkel, E., Lipschutz-Broch, L., Lieberman, J.A. (1992). Antipsychotic medication effects on neuropsychological functions. *Psychopharmacology bulletin* 28:353-366.

Bora, E., Lin, A., Wood, S.J., Yung, A.R., McGorry, P.D., Pantelis, C. (2014). Cognitive deficits in youth with familial and clinical high risk to psychosis: a systematic review and meta-analysis. *Acta Psychiatrica Scandinavica* 130:1-15.

Brand, N., Jolles, J. (1985). Learning and retrieval rate of words presented auditorily and visually. *The Journal of general psychology* 112:201-210.

Buckley, P.F., Miller, B.J., Lehrer, D.S., Castle, D.J. (2009). Psychiatric comorbidities and schizophrenia. *Schizophrenia bulletin* 35:383-402.

Chua, S.E., Cheung, C., Cheung, V., Tsang, J.T., Chen, E.Y., Wong, J.C., Cheung, J.P., Yip, L., Tai, K.S., Suckling, J., McAlonan, G.M. (2007). Cerebral grey, white matter and csf in never-medicated, first-episode schizophrenia. *Schizophrenia research* 89:12-21.

Chung, Y.S., Kang, D.H., Shin, N.Y., Yoo, S.Y., Kwon, J.S. (2008). Deficit of theory of mind in individuals at ultra-high-risk for schizophrenia. *Schizophrenia research* 99:111-118.

Crossley, N.A., Mechelli, A., Fusar-Poli, P., Broome, M.R., Matthiasson, P., Johns, L.C., Bramon, E., Valmaggia, L., Williams, S.C., McGuire, P.K. (2009). Superior temporal lobe dysfunction and frontotemporal dysconnectivity in subjects at risk of psychosis and in first-episode psychosis. *Human brain mapping* 30:4129-4137.

Cullen, A.E., Zunszain, P.A., Dickson, H., Roberts, R.E., Fisher, H.L., Pariante, C.M., Laurens, K.R. (2014). Cortisol awakening response and diurnal cortisol among children at elevated risk for schizophrenia: relationship to psychosocial stress and cognition. *Psychoneuroendocrinology* 46:1-13.

Davis, L.L., Broome, M.E., Cox, R.P. (2002). Maximizing retention in community-based clinical trials. *Journal of nursing scholarship : an official publication of Sigma Theta Tau International Honor Society of Nursing* 34:47-53.

de Achaval, D., Costanzo, E.Y., Villarreal, M., Jauregui, I.O., Chiodi, A., Castro, M.N., Fahrer, R.D., Leiguarda, R.C., Chu, E.M., Guinjoan, S.M. (2010). Emotion processing and theory of mind in schizophrenia patients and their unaffected first-degree relatives. *Neuropsychologia* 48:1209-1215.

Doody, G.A., Gotz, M., Johnstone, E.C., Frith, C.D., Owens, D.G. (1998). Theory of mind and psychoses. *Psychological medicine* 28:397-405.

Elvevag, B., Goldberg, T.E. (2000). Cognitive impairment in schizophrenia is the core of the disorder. *Critical reviews in neurobiology* 14:1-21.

Frommann, I., Pukrop, R., Brinkmeyer, J., Bechdolf, A., Ruhrmann, S., Berning, J., Decker, P., Riedel, M., Moller, H.J., Wolwer, W., Gaebel, W., Klosterkotter, J., Maier, W., Wagner, M. (2011). Neuropsychological profiles in different at-risk states of psychosis: executive control impairment in the early--and additional memory dysfunction in the late--prodromal state. *Schizophrenia bulletin* 37:861-873.

Fusar-Poli, P., Deste, G., Smieskova, R., Barlati, S., Yung, A.R., Howes, O., Stieglitz, R.D., Vita, A., McGuire, P., Borgwardt, S. (2012). Cognitive functioning in prodromal psychosis: a meta-analysis. *Archives of General Psychiatry* 69:562-571.

Garety, P.A., Bebbington, P., Fowler, D., Freeman, D., Kuipers, E. (2007). Implications for neurobiological research of cognitive models of psychosis: a theoretical paper. *Psychological medicine* 37:1377-1391.

Gottesman, I.I., Wolfgram, D.L. (1991). *Schizophrenia genesis : The origins of madness.* : New York: Freeman.

Green, M.F. (1996). What are the functional consequences of neurocognitive deficits in schizophrenia?. *The American Journal of Psychiatry* 153:321-330.

Green, M.F., Kern, R.S., Braff, D.L., Mintz, J. (2000). Neurocognitive deficits and functional outcome in schizophrenia: are we measuring the "right stuff"?. *Schizophrenia bulletin* 26:119-136.

Green, M.F., Kern, R.S., Heaton, R.K. (2004). Longitudinal studies of cognition and functional outcome in schizophrenia: implications for MATRICS. *Schizophrenia research* 72:41-51.

Gur, R.E., Turetsky, B.I., Cowell, P.E., Finkelman, C., Maany, V., Grossman, R.I., Arnold, S.E., Bilker, W.B., Gur, R.C. (2000). Temporolimbic volume reductions in schizophrenia. *Archives of General Psychiatry* 57:769-775.

Hansen, W.B., Collins, L.M., Malotte, C.K., Johnson, C.A., Fielding, J.E. (1985). Attrition in prevention research. *Journal of Behavioral Medicine* 8:261-275.

Heinrichs, R.W., Zakzanis, K.K. (1998). Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. *Neuropsychology* 12:426-445.

Hirayasu, Y., Tanaka, S., Shenton, M.E., Salisbury, D.F., DeSantis, M.A., Levitt, J.J., Wible, C., Yurgelun-Todd, D., Kikinis, R., Jolesz, F.A., McCarley, R.W. (2001). Prefrontal gray matter volume reduction in first episode schizophrenia. *Cerebral cortex (New York, N.Y.: 1991)* 11:374-381.

Horton, N.J., Kleinman, K.P. (2007). Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *The American statistician* 61:79-90.

Howie, E.K., Straker, L.M. (2016). Rates of attrition, non-compliance and missingness in randomized controlled trials of child physical activity interventions using accelerometers: A brief methodological review. *Journal of Science and Medicine in Sport* 19:830-836.

Hughes, C., Kumari, V., Das, M., Zachariah, E., Ettinger, U., Sumich, A., Sharma, T. (2005). Cognitive functioning in siblings discordant for schizophrenia. *Acta Psychiatrica Scandinavica* 111:185-192.

Irani, F., Seligman, S., Kamath, V., Kohler, C., Gur, R.C. (2012). A meta-analysis of emotion perception and functional outcomes in schizophrenia. *Schizophrenia research* 137:203-211.

Keefe, R.S., Perkins, D.O., Gu, H., Zipursky, R.B., Christensen, B.K., Lieberman, J.A. (2006). A longitudinal study of neurocognitive function in individuals at-risk for psychosis. *Schizophrenia research* 88:26-35.

Kim, H.S., Shin, N.Y., Choi, J.S., Jung, M.H., Jang, J.H., Kang, D.H., Kwon, J.S. (2010). Processing of facial configuration in individuals at ultra-high risk for schizophrenia. *Schizophrenia research* 118:81-87.

Kim, K.R., Park, J.Y., Song, D.H., Koo, H.K., An, S.K. (2011). Neurocognitive performance in subjects at ultrahigh risk for schizophrenia: a comparison with first-episode schizophrenia. *Comprehensive psychiatry* 52:33-40.

Klemm, S., Schmidt, B., Knappe, S., Blanz, B. (2006). Impaired working speed and executive functions as frontal lobe dysfunctions in young first-degree relatives of schizophrenic patients. *European child & adolescent psychiatry* 15:400-408.

Konings, M., Bak, M., Hanssen, M., van Os, J., Krabbendam, L. (2006). Validity and reliability of the CAPE: a self-report instrument for the measurement of psychotic experiences in the general population. *Acta Psychiatrica Scandinavica* 114:55-61.

Korver, N., Quee, P.J., Boos, H.B.M., Simons, C.J.P., de Haan, L., GPOUPinvestigators, Investigators, G. (2012). Genetic Risk and Outcome of Psychosis (GROUP), a multi-site longitudinal cohort study focused on gene-environment interaction: objectives, sample characteristics, recruitment and assessment methods. *International Journal of Methods in Psychiatric Research* 21:205-221.

Kremen, W.S., Seidman, L.J., Pepple, J.R., Lyons, M.J., Tsuang, M.T., Faraone, S.V. (1994). Neuropsychological risk indicators for schizophrenia: a review of family studies. *Schizophrenia bulletin* 20:103-119.

Lachenbruch, P. (1976). Analysis of data with clumping at zero. 18:351-356.

Lachenbruch, P. (1992). *Utility of regression analysis in epidemiologic studies of the elderly.*: Oxford University Press.

Laurent, A., Duly, D., Murry, P., Foussard, N., Boccara, S., Mingat, F., Dalery, J., d'Amato, T. (2001). WCST performance and schizotypal features in the first-degree relatives of patients with schizophrenia. *Psychiatry research* 104:133-144.

Lencz, T., Smith, C.W., McLaughlin, D., Auther, A., Nakayama, E., Hovey, L., Cornblatt, B.A. (2006). Generalized and specific neurocognitive deficits in prodromal schizophrenia. *Biological psychiatry* 59:863-871.

Little, R.J.A., Rubin, D.B. (2002). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.

Meijer, J., Simons, C.J., Quee, P.J., Verweij, K., GROUP Investigators. (2012). Cognitive alterations in patients with non-affective psychotic disorder and their unaffected siblings and parents. *Acta Psychiatrica Scandinavica* 125:66-76.

Meiran, N., Levine, J., Meiran, N., Henik, A. (2000). Task set switching in schizophrenia. *Neuropsychology* 14:471-482.

Molenberghs, G., Verbeke, G. (2006). *Models for Discrete Longitudinal Data*. New York: Springer-Verlag.

Myles-Worsley, M., Ord, L.M., Ngiralmau, H., Weaver, S., Blailes, F., Faraone, S.V. (2007). The Palau Early Psychosis Study: neurocognitive functioning in high-risk adolescents. *Schizophrenia research* 89:299-307.

Nuechterlein, K.H., Barch, D.M., Gold, J.M., Goldberg, T.E., Green, M.F., Heaton, R.K. (2004). Identification of separable cognitive factors in schizophrenia. *Schizophrenia research* 72:29-39.

Nuechterlein, K.H., Dawson, M.E. (1984). Information processing and attentional functioning in the developmental course of schizophrenic disorders. *Schizophrenia bulletin* 10:160-203.

Pijnenborg, G.H., Spikman, J.M., Jeronimus, B.F., Aleman, A. (2013). Insight in schizophrenia: associations with empathy. *European archives of psychiatry and clinical neuroscience* 263:299-307.

Pijnenborg, G.H., Withaar, F.K., Evans, J.J., van den Bosch, R.J., Timmerman, M.E., Brouwer, W.H. (2009). The predictive value of measures of social cognition for community functioning in schizophrenia: implications for neuropsychological assessment. *Journal of the International Neuropsychological Society : JINS* 15:239-247.

Pukrop, R., Ruhrmann, S., Schultze-Lutter, F., Bechdolf, A., Brockhaus-Dumke, A., Klosterkotter, J. (2007). Neurocognitive indicators for a conversion to psychosis: comparison of patients in a potentially initial prodromal state who did or did not convert to a psychosis. *Schizophrenia research* 92:116-125.

Quee, P.J., Alizadeh, B.Z., Aleman, A., van den Heuvel, E.R., GROUP Investigators. (2014). Cognitive subtypes in non-affected siblings of schizophrenia patients: characteristics and profile congruency with affected family members. *Psychological medicine* 44:395-405.

Riecher-Rossler, A., Pflueger, M.O., Aston, J., Borgwardt, S.J., Brewer, W.J., Gschwandtner, U., Stieglitz, R.D. (2009). Efficacy of using cognitive status in predicting psychosis: a 7-year follow-up. *Biological psychiatry* 66:1023-1030.

Robertson, J. S., Bolinger, K., Glasser, L. M., Sloane, N. J., and Gross, R. (1996). Chapter 1. In *CRC standard mathematical tables and formulae*, Zwillinger, D. (ed), 71. Boca Raton: CRC Press.

Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys*. New York: J. Wiley & Sons.

Seidman, L.J., Pousada-Casal, A., Scala, S., Meyer, E.C., Stone, W.S., Thermenos, H.W., Molokotos, E., Agnew-Blais, J., Tsuang, M.T., Faraone, S.V. (2016). Auditory Vigilance and Working Memory in Youth at Familial Risk for Schizophrenia or Affective Psychosis in the Harvard Adolescent Family High Risk Study. *Journal of the International Neuropsychological Society : JINS* 22:1026-1037.

Snitz, B.E., Macdonald, A.W.,3rd, Carter, C.S. (2006). Cognitive deficits in unaffected first-degree relatives of schizophrenia patients: a meta-analytic review of putative endophenotypes. *Schizophrenia bulletin* 32:179-194.

Stefanis, N.C., Hanssen, M., Smirnis, N.K., Avramopoulos, D.A., Evdokimidis, I.K., Stefanis, C.N., Verdoux, H., Van Os, J. (2002). Evidence that three dimensions of psychosis have a distribution in the general population. *Psychological medicine* 32:347-358.

Thompson, W.K., Savla, G.N., Vahia, I.V., Depp, C.A., Oâ€™Hara, R., Jeste, D.V., Palmer, B.W. (2012). Characterizing Trajectories of Cognitive Functioning in Older Adults with Schizophrenia: Does Method Matter?. *Schizophrenia research* 143:90-96.

Tooze, J.A., Grunwald, G.K., Jones, R.H. (2002). Analysis of repeated measures data with clumping at zero. *Statistical methods in medical research* 11:341-355.

van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical methods in medical research* 16:219-242.

van der Meer, F.J., Meijer, J.H., Meijer, C.J., van den Brink, W., Velthorst, E., Genetic Risk and Outcome of Psychosis (GROUP) Investigators. (2014). Cognitive functioning associated with stimulant use in patients with non-affective psychosis, their unaffected siblings and healthy controls. *Psychological medicine* 44:1901-1911.

van Donkersgoed, R.J., Wunderink, L., Nieboer, R., Aleman, A., Pijnenborg, G.H. (2015). Social Cognition in Individuals at Ultra-High Risk for Psychosis: A Meta-Analysis. *PloS one* 10:e0141075.

van Os, J., Linscott, R.J., Myin-Germeys, I., Delespaul, P., Krabbendam, L. (2009). A systematic review and meta-analysis of the psychosis continuum: evidence for a psychosis proneness-persistence-impairment model of psychotic disorder. *Psychological medicine* 39:179-195.

van Rijn, S., Schothorst, P., Wout, M., Sprong, M., Ziermans, T., van Engeland, H., Aleman, A., Swaab, H. (2011). Affective dysfunctions in adolescents at risk for psychosis: emotion awareness and social functioning. *Psychiatry research* 187:100-105.

van 't Wout, M., Aleman, A., Kessels, R.P., Laroi, F., Kahn, R.S. (2004). Emotional processing in a non-clinical psychosis-prone sample. *Schizophrenia research* 68:271-281.

van Winkel, R., GROUP Investigators. (2015). Further Evidence That Cannabis Moderates Familial Correlation of Psychosis-Related Experiences. *PloS one* 10:e0137625.

Versmissen, D., Janssen, I., Myin-Germeys, I., Mengelers, R., Campo, J.A., van Os, J., Krabbendam, L. (2008). Evidence for a relationship between mentalising deficits and paranoia over the psychosis continuum. *Schizophrenia research* 99:103-110.

Wechsler, D. (1997). *WAIS-III: Wechsler adult intelligence scale, 3rd edn. administration and scoring manual.* San Antonio, TX, USA: Psychological Corporation.

WHO. (1990). Composite international diagnostic interview (CIDI):(a) CIDI-interview (version 1.0), (b) CIDI-user manual, (c) CIDI-training manual (d) CIDI-computer programs.

Wigman, J.T., Vollebergh, W.A., Raaijmakers, Q.A., Iedema, J., van Dorsselaer, S., Ormel, J., Verhulst, F.C., van Os, J. (2011). The structure of the extended psychosis phenotype in early adolescence--a cross-sample replication. *Schizophrenia bulletin* 37:850-860.

Wing, J.K., Babor, T., Brugha, T., Burke, J., Cooper, J.E., Giel, R., Jablenski, A., Regier, D., Sartorius, N. (1990). SCAN. Schedules for Clinical Assessment in Neuropsychiatry. *Archives of General Psychiatry* 47:589-593.

Yung, A.R., Phillips, L.J., Yuen, H.P., Francey, S.M., McFarlane, C.A., Hallgren, M., McGorry, P.D. (2003). Psychosis prediction: 12-month follow up of a high-risk ("prodromal") group. *Schizophrenia research* 60:21-32.

Zeger, S.L., Karim, M.R. (1991). Generalized Linear Models With Random Effects; A Gibbs Sampling Approach. *Journal of the American Statistical Association* 86:79-86.

Supplementary Materials

Supplementary Methods

The general form of the mixture distribution model where the outcome has lots of zeros as defined as

$$f(y) = \begin{cases} P(Y=0) & \text{if } y = 0 \\ [1 - P(Y=0)]d(y) & \text{if } y > 0 \end{cases}$$

where *d(y)* is the probability density defined when *y* > 0 (Lachenbruch, 1976; Lachenbruch, 1992; Tooze et al., 2002).

Let Y_{hc} be score of the frequency of psychotic experiences or the amount of distress of those experiences for sibling *h* (=1, 2, …, *m*) at time *c* (=1, 2, …,n_h), and R_{hc} represent the occurrence of frequency/distress variable where $R_{hc} = \begin{cases} 0, & \text{if } Y_{hc} = 0 \\ 1, & \text{if } Y_{hc} > 0 \end{cases}$ and it has conditional probabilities

$$P(R_{hc} = r_{hc}|\theta_1) = \begin{cases} 1 - p_{hc}(\theta_1), & if\ r_{hc} = 0 \\ p_{hc}(\theta_1), & if\ r_{hc} = 1 \end{cases}$$

where $\boldsymbol{\theta}_1 = \left[\boldsymbol{\beta}_1^{/}, u_{1h}\right]^{/}$ is a vector of fixed occurrence effects $\boldsymbol{\beta}_1$, and random intercept unit (sibling) occurrence effect u_{1h}. We assume a logistic model for occurrence of the frequency of psychotic experiences or the amount of distress of those experiences so that $logit\left(p_{hc}(\theta_1)\right) = \boldsymbol{X}_{1hc}^{/}\boldsymbol{\beta}_1 + u_{1h}$ where X_{1hc} is a vector of independent variables for occurrence. Let $S_{hc} \equiv [Y_{hc}|R_{hc} = 1]$ be the intensity variable of frequency/distress psychotic experiences with pdf $f(s_{hc}|\boldsymbol{\theta}_2)$ for $s_{hc} > 0$, where $\boldsymbol{\theta}_2 = \left[\boldsymbol{\beta}_2^{/}, u_{2h}\right]^{/}$ is a vector of fixed effects of $\boldsymbol{\beta}_2$, and random intercept unit (sibling) of intensity effect u_{2h}. We assume a lognormal model for intensity so that $log(S_{hc}|\boldsymbol{\theta}_2) \sim N\left(\boldsymbol{X}_{2hc}^{/}\boldsymbol{\beta}_2 + u_{2h}, \sigma_e^2\right)$ where $\boldsymbol{X}_{2hc}$ is a vector of independent variables for intensity. We allow the random intercepts for both occurrence and intensity of the frequency of psychotic experiences or the amount of distress of those experiences with the assumption that they are bivariate normal and to be correlated as

$\begin{bmatrix} u_{1hc} \\ u_{2hc} \end{bmatrix} \sim BVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right)$, where ρ is the correlation coefficient.

Based on this assumption, the subject-specific average intensity of the frequency of psychotic experiences or the amount of distress of those experiences is $E(S_{hc}|\boldsymbol{\theta}_2) = exp\left(\boldsymbol{X}_{2hc}^{/}\boldsymbol{\beta}_2 + u_{2h} + \sigma_e^2/2\right)$ and the marginal average intensity is $E(S_{hc}|\boldsymbol{\theta}_2) = exp\left(\boldsymbol{X}_{2hc}^{/}\boldsymbol{\beta}_2 + \sigma_2^2/2 + \sigma_e^2/2\right)$.

The pdf of Y_{hc} is $f(y_{hc}|\boldsymbol{\theta}) = P(R_{hc} = 0|\boldsymbol{\theta}_1)\delta_0(y_{hc}) + P(R_{hc} = 1|\boldsymbol{\theta}_1)f(s_{hc}|\boldsymbol{\theta}_2)$

$$= [1 - p_{hc}(\boldsymbol{\theta}_1)]\delta_0(y_{hc}) + p_{hc}(\boldsymbol{\theta}_1)f(s_{hc}|\boldsymbol{\theta}_2)$$

Where $\boldsymbol{\theta} = \left[\boldsymbol{\theta}_1^{/}, \boldsymbol{\theta}_2^{/}\right]$ and $\delta_0(y_{hc})$ is Dirac delta function (Robertson et al., 1996) such that

$$\begin{cases} \int_{-\infty}^{\infty} \delta_0(y)dy_{hc} = 1 \\ \delta_0(y) = 0\ when\ y_{hc} \neq 0 \end{cases}$$

Therefore, the likelihood function for the correlated mixed distribution of Y_{hc} is

$$L(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma_1, \sigma_2, \sigma_e, \rho)$$
$$= \prod_{h=1}^{m} \int_{u_{1h}} \int_{u_{2h}} \prod_{c=1}^{n_h} f(y_{hc}|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, u_{1h}, u_{2h}) f(u_{1h}, u_{2h}|\sigma_1, \sigma_2, \sigma_e, \rho)\, du_{1h} du_{2h}$$
$$= \prod_{h=1}^{m} \int_{u_{1h}} \int_{u_{2h}} \prod_{c=1}^{n_h} [1 - p_{hc}(\boldsymbol{\beta}_1, u_{1h})]^{1-r_{hc}} [p_{hc}(\boldsymbol{\beta}_1, u_{1h})]^{r_{hc}}$$
$$\times f(s_{hc}|\boldsymbol{\beta}_2, u_{2h}) f(u_{1h}, u_{2h}|\sigma_1, \sigma_2, \sigma_e, \rho) du_{1h} du_{2h} \ldots\ldots\ldots\ldots (1)$$

If u_{1h} and u_{2h} are independent i.e. $\rho = 0$, the equation (1) will be decomposed into two parts that correspond to occurrence and intensity parts and termed as uncorrelated mixed distribution as

$$L(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma_1, \sigma_2, \sigma_e)$$
$$= \prod_{h=}^{m} \int_{u_{1h}} \prod_{c=1}^{n_h} [1 - p_{hc}(\boldsymbol{\theta}_1)]^{1-r_{hc}} [p_{hc}(\boldsymbol{\theta}_1)]^{r_{hc}} f(u_{1h}|\sigma_1) du_{1h}$$
$$\times \prod_{h=1}^{m} \int_{u_{2h}} \prod_{c=1}^{n_h} f(s_{hc}|\boldsymbol{\theta}_2) f(u_{2h}|\sigma_2, \sigma_e) du_{2h} \ldots\ldots\ldots\ldots (2)$$

The correlated mixed-distribution model (1) is maximized by using quasi-Newton optimization of a likelihood approximated by adaptive Gaussian quadrature (Molenberghs and Verbeke, 2006; Zeger and Karim, 1991; Tooze et al., 2002). Two parts of the uncorrelated mixed distribution model (2) are maximized separately by adaptive Gaussian quadrature. A SAS macro MIXCORR developed by Tooze et al (Tooze et al., 2002) is used to fit the correlated and uncorrelated model. In short, within the MIXCORR macro, we fit the models with PROC GENMOD and PROC NLMIXED and the results of parameter estimates are used as starting values for the final estimation of the model parameters for both correlated and uncorrelated mixed effects models using PROC NLMIXED. The details of the estimation procedures have been found in elsewhere (Tooze et al., 2002).

Supplementary Table S1: Differences at baseline as a function of follow-up attrition of all characteristics for controls (N=386)

Variables	Type	CAPE[a] Positive symptoms, frequency				CAPE[a] Positive symptoms, distress			
		Mean or %	SD	N	P-value	Mean or %	SD	N	P-value
Age	No follow-up	22.70	4.77	108	**0.034**	22.79	4.77	110	0.054
	Follow-up	24.24	5.71	278		24.21	5.73	276	
Gender, Female	No follow-up	48.15	...	52	0.305	47.27	...	52	0.209
	Follow-up	53.96	...	150		54.35	...	150	
Education[b]	No follow-up	4.93	1.88	108	0.141	4.96	1.89	110	0.221
	Follow-up	5.26	1.74	277		5.24	1.74	275	
Ethnicity, Caucasian	No follow-up	88.35	...	91	0.875	87.62	...	92	0.829
	Follow-up	90.37	...	244		90.67	...	243	
Cannabis use last 12 months, Yes	No follow-up	23.36	...	25	0.419	22.94	...	25	0.492
	Follow-up	19.64	...	54		19.78	...	54	
Neurocognition:									
Immediate Recall[c]	No follow-up	29.16	4.66	104	0.767	29.07	4.68	106	0.971
	Follow-up	29.14	5.19	276		29.18	5.18	274	
Retention rate[d]	No follow-up	0.83	0.15	103	0.554	0.83	0.15	105	0.595
	Follow-up	0.84	0.15	271		0.84	0.15	269	
Sensitivity Index[e]	No follow-up	98.25	3.82	102	0.622	98.11	4.10	104	0.745
	Follow-up	97.62	7.76	249		97.67	7.73	247	
Accuracy Cost Score[f]	No follow-up	0.13	0.20	96	0.522	0.12	0.20	98	0.702
	Follow-up	0.12	0.19	253		0.12	0.19	251	
IQ[g]	No follow-up	106.02	14.21	105	**0.014**	105.81	14.25	107	**0.008**
	Follow-up	110.63	14.47	275		110.74	14.42	273	
Block Design[h]	No follow-up	46.10	13.43	105	**0.013**	45.66	13.75	107	**0.005**
	Follow-up	49.63	13.45	276		49.83	13.27	274	
Digit Symbol-coding[i]	No follow-up	85.01	13.22	105	0.751	84.76	13.22	107	0.568
	Follow-up	85.16	14.29	276		85.26	14.29	274	
Information[j]	No follow-up	17.57	5.21	105	**0.033**	17.63	5.18	107	**0.042**
	Follow-up	18.92	4.55	276		18.91	4.56	274	
Arithmetic[k]	No follow-up	14.16	4.29	105	**0.004**	14.17	4.32	107	**0.004**
	Follow-up	15.53	4.13	276		15.54	4.12	274	
Benton Facial[l]	No follow-up	23.06	2.03	104	0.162	22.99	2.08	106	0.083
	Follow-up	23.30	2.01	274		23.33	1.98	272	

Supplementary Table S1-Continued

Variables	Type	CAPE[a] Positive symptoms, frequency				CAPE[a] Positive symptoms, distress			
		Mean or %	SD	N	P-value	Mean or %	SD	N	P-value
Social Cognition:									
Hinting Task	No follow-up	19.04	1.20	103	0.480	19.03	1.21	105	0.433
	Follow-up	19.13	1.21	273		19.13	1.20	271	
Degraded Facial Affect Recognition									
Percent happy faces	No follow-up	90.13	10.40	100	0.131	90.01	10.34	102	0.181
	Follow-up	88.33	10.77	256		88.36	10.81	254	
Percent fearful faces	No follow-up	56.75	18.48	100	0.658	56.50	18.40	102	0.805
	Follow-up	55.74	17.13	256		55.83	17.15	254	
Percent angry faces	No follow-up	71.75	18.37	100	0.906	71.51	18.31	102	0.946
	Follow-up	71.44	19.01	256		71.53	19.03	254	
Percent neutral faces	No follow-up	82.63	13.95	100	0.607	82.60	13.92	102	0.625
	Follow-up	81.32	15.22	256		81.32	15.24	254	
Outcome variables:									
CAPE: Frequency of positive symptoms	No follow-up	0.24	0.19	99	**0.011**	...	...	...	...
	Follow-up	0.20	0.18	265		...	...	...	
CAPE: Distress of positive symptoms	No follow-up	...	...	...	...	0.46	0.47	99	0.105
	Follow-up	...	...	...		0.37	0.43	265	

[a]CAPE: Community Assessment of Psychic Experiences; [b]Education (Verhage): range 0 (no education), 3-5 (school diploma), 6-8 (professional education/university degree); [c]Immediate Recall: Word Learning Task (WLT) Immediate Recall; [d]Retention rate: WLT Retention Rate; [e]Sensitivity Index: Continuous Performance Test Sensitivity Index; [f]Accuracy Cost Score: Response Shifting Task Accuracy Cost Score; [g]IQ: WAIS-III Intelligence Quotient; [h]Block Design: WAIS-III Block Design; [i]Digit Symbol Coding: WAIS-III Digit Symbol Coding; [j]Information: WAIS-III Information; [k]Arithmetic: WAIS-III Arithmetic; [l]Benton Facial: Benton Facial Recognition Test.

Supplementary Table S2: Differences at baseline as a function of follow-up attrition of all characteristics for siblings (N=873)

Variables	Type	CAPE[a] Positive symptoms, frequency				CAPE[a] Positive symptoms, distress			
		Mean or %	SD	N	P-value	Mean or %	SD	N	P-value
Age	No follow-up	24.86	5.43	222	0.755	24.80	5.45	224	0.603
	Follow-up	24.94	5.30	651		24.96	5.29	649	
Gender, Female	No follow-up	48.20	...	107	**0.035**	47.77	...	107	**0.023**
	Follow-up	56.3	...	367		56.55	...	367	
Education[b]	No follow-up	4.50	2.15	215	**0.001**	4.47	2.15	217	**<0.001**
	Follow-up	5.10	2.04	642		5.12	2.04	640	
Ethnicity, Caucasian	No follow-up	69.37	...	154	**<0.001**	69.64	...	156	**<0.001**
	Follow-up	87.04	...	564		87.00	...	562	
Cannabis use last 12 months, Yes	No follow-up	29.36	...	64	**0.001**	29.09	...	64	**0.001**
	Follow-up	18.52	...	120		18.58		120	
Neurocognition:									
Immediate Recall[c]	No follow-up	26.11	5.83	210	**0.001**	26.17	5.84	212	**0.002**
	Follow-up	27.54	5.51	638		27.53	5.51	636	
Retention rate[d]	No follow-up	0.82	0.17	207	**0.020**	0.82	0.17	209	**0.010**
	Follow-up	0.85	0.17	630		0.85	0.17	628	
Sensitivity Index[e]	No follow-up	96.26	11.62	193	0.353	96.26	11.54	196	0.279
	Follow-up	96.87	9.96	595		96.88	9.99	592	
Accuracy Cost Score[f]	No follow-up	0.14	0.21	196	0.876	0.13	0.21	199	0.902
	Follow-up	0.14	0.22	591		0.14	0.22	588	
IQ[g]	No follow-up	97.75	13.83	204	**<0.001**	97.87	13.93	206	**<0.001**
	Follow-up	104.23	15.73	636		104.21	15.72	634	
Block Design[h]	No follow-up	41.78	15.74	211	**<0.001**	41.69	15.82	213	**<0.001**
	Follow-up	46.72	14.40	641		46.77	14.35	639	
Digit Symbol-coding[i]	No follow-up	76.92	15.00	213	**0.001**	76.98	14.98	215	**0.001**
	Follow-up	81.19	15.27	642		81.19	15.29	640	
Information[j]	No follow-up	15.43	4.97	212	**<0.001**	15.52	4.98	214	**<0.001**
	Follow-up	17.16	5.25	642		17.13	5.26	640	
Arithmetic[k]	No follow-up	12.64	4.56	213	**<0.001**	12.66	4.55	215	**<0.001**
	Follow-up	14.17	4.37	641		14.17	4.37	639	
Benton Facial[l]	No follow-up	23.40	2.02	209	0.223	23.38	2.01	211	0.271
	Follow-up	23.15	2.21	638		23.15	2.21	636	

Supplementary Table S2-continued

Variables	Type	CAPE[a] Positive symptoms, frequency				CAPE[a] Positive symptoms, distress			
		Mean or %	SD	N	P-value	Mean or %	SD	N	P-value
Social Cognition:									
Hinting Task	No follow-up	18.81	1.69	212	0.901	18.81	1.69	214	0.900
	Follow-up	18.80	1.70	637		18.80	1.71	635	
Degraded Facial Affect Recognition (DFAR)									
Percent happy faces	No follow-up	88.87	11.18	201	0.338	88.76	11.16	204	0.456
	Follow-up	88.24	10.47	606		88.28	10.48	603	
Percent fearful faces	No follow-up	54.91	19.83	201	0.701	54.84	19.72	204	0.749
	Follow-up	54.39	18.99	606		54.42	19.02	603	
Percent angry faces	No follow-up	69.34	17.99	201	0.860	69.27	17.97	204	0.807
	Follow-up	69.45	19.31	606		69.48	19.32	603	
Percent neutral faces	No follow-up	81.19	15.23	201	0.219	81.43	15.26	204	0.121
	Follow-up	79.91	15.13	606		79.82	15.11	603	
Outcome variables:									
CAPE: PE frequency	No follow-up	0.22	0.25	183	0.159	...	...	...	...
	Follow-up	0.22	0.19	574		...	...	...	
CAPE: PE Distress	No follow-up	...	...	...	...	0.37	0.47	182	0.126
	Follow-up	...	...	...		0.42	0.47	566	

[a]CAPE: Community Assessment of Psychic Experiences; [b]Education (Verhage): range 0 (no education), 3-5 (school diploma), 6-8 (professional education/university degree); [c]Immediate Recall: Word Learning Task (WLT) Immediate Recall; [d]Retention rate: WLT Retention Rate; [e]Sensitivity Index: Continuous Performance Test Sensitivity Index; [f]Accuracy Cost Score: Response Shifting Task Accuracy Cost Score; [g]IQ: WAIS-III Intelligence Quotient; [h]Block Design: WAIS-III Block Design; [i]Digit Symbol Coding: WAIS-III Digit Symbol Coding; [j]Information: WAIS-III Information; [k]Arithmetic: WAIS-III Arithmetic; [l]Benton Facial: Benton Facial Recognition Test.

Supplementary Table S3: Spearman correlation coefficients of all neuro and social cognition measurements for healthy controls (n = 318)[a]

	WLTTIR	WLTRR	CPT_SEN	RST_AC	IQ	WAIBD	WAIDS	WAIIN	WAICA	BENTFR	HINTS	DFR_PHF	DFR_PFF	DFR_PAF
WLTRR	0.36***													
CPT_SEN	0.12*	0.08												
RST_AC	-0.05	0.02	0.09											
IQ	0.27***	0.11*	0.09	-0.16**										
WAIBD	0.15**	0.03	0.05	-0.16**	0.65***									
WAIDS	0.28***	0.18***	0.11*	-0.06	0.59***	0.23***								
WAIIN	0.12*	-0.02	0.05	-0.12*	0.68***	0.30***	0.16**							
WAICA	0.20***	0.07	0.02	-0.1	0.75***	0.36***	0.22***	0.56***						
BENTFR	0.17**	0.13*	0.09	0.05	0.11	0.07	0.15**	0.14*	0					
HINTS	0.1	0.16**	0.13*	0.06	0.13*	0.05	0.09	0.07	0.1	0.15**				
DFR_PHF	0.14*	0.11	0.06	-0.07	0.01	0.02	0.04	0	-0.02	0.13*	0.1			
DFR_PFF	0.16**	0.05	0.09	-0.02	0.15**	0.14*	0.18**	0	0.04	0.01	0.02	0.18**		
DFR_PAF	0.09	-0.05	0.01	0	-0.01	0.03	0.03	-0.04	-0.05	0.11*	0.06	0.15**	0.28***	
DFR_PNF	0.03	-0.02	0.08	-0.07	0.04	0.08	0.02	0.07	0.05	0.16**	0.04	-0.05	-0.13*	-0.11*

[a]The table represents Spearman's correlation coefficients and their significance level; WLTTIR: Word Learning Task (WLT) Immediate Recall; WLTRR: WLT Retention Rate; CPT_SEN: Continuous Performance Test Sensitivity Index; RST_AC: Response Shifting Task Accuracy Cost Score; IQ: WAIS-III Intelligence Quotient; WAIBD: WAIS-III Block Design; WAIDS: WAIS-III Digit Symbol Coding; WAIIN: WAIS-III Information; WAICA: WAIS-III Arithmetic; BENTFR: Benton Facial Recognition Test; Hint: Hinting Task; DFR_PHF: Degraded Facial Affect Recognition (DFAR) Percent happy faces; DFR_PFF: DFAR Percent fearful faces; DFR_PAF: DFAR Percent angry faces; DFR_PNF: DFAR Percent neutral faces; Significance levels: ***P <0 .001, **P <0 .01 and *P < 0.05.

Supplementary Table S4: Spearman correlation coefficients of all neuro and social cognition measurements for siblings (n = 709)[a]

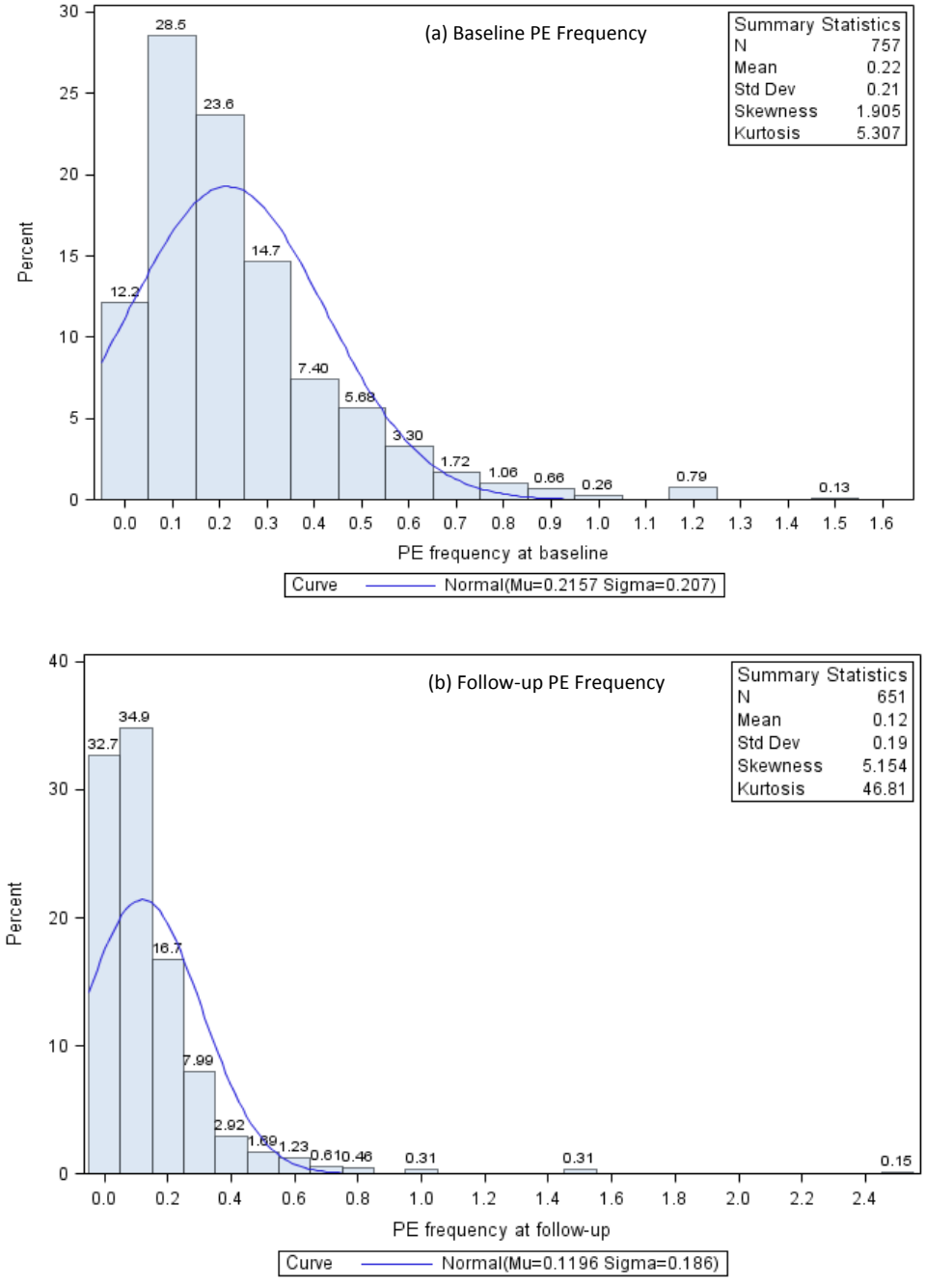
	WLTTIR	WLTRR	CPT_SEN	RST_AC	IQ	WAIBD	WAIDS	WAIIN	WAICA	BENTFR	HINTS	DFR_PHF	DFR_PFF	DFR_PAF
WLTRR	0.27***													
CPT_SEN	0.17***	0.04												
RST_AC	-0.08*	-0.07	-0.05											
IQ	0.32***	0.10**	0.30***	-0.13***										
WAIBD	0.19***	0.05	0.24***	-0.07	0.74***									
WAIDS	0.29***	0.14***	0.21***	-0.10**	0.64***	0.34***								
WAIIN	0.25***	0.07*	0.23***	-0.11**	0.77***	0.43***	0.30***							
WAICA	0.23***	0.04	0.25***	-0.10**	0.82***	0.50***	0.37***	0.63***						
BENTFR	0.10**	0.09*	0.05	-0.05	0.13***	0.11**	0.08*	0.13***	0.12**					
HINTS	0.17***	0.02	0.06	-0.05	0.20***	0.12***	0.15***	0.20***	0.19***	0.05				
DFR_PHF	0.06	0.06	0.15***	-0.06	0.14***	0.13***	0.14***	0.12**	0.07	0.19***	0.05			
DFR_PFF	0.10**	0.06	0.11**	-0.08*	0.06	0.08*	0.05	0	0.04	0.10*	0.01	0.15***		
DFR_PAF	0.12**	0.11**	0.06	-0.04	0.02	0.08*	0.08*	-0.04	-0.04	0.09*	0.04	0.14***	0.27***	
DFR_PNF	0.08*	-0.06	0.01	-0.07	0.11**	0.05	0.07*	0.11**	0.11**	0.11**	0.03	0	-0.04	-0.04

[a]The table represents Spearman's correlation coefficients and their significance level; WLTTIR: Word Learning Task (WLT) Immediate Recall; WLTRR: WLT Retention Rate; CPT_SEN: Continuous Performance Test Sensitivity Index; RST_AC: Response Shifting Task Accuracy Cost Score; IQ: WAIS-III Intelligence Quotient; WAIBD: WAIS-III Block Design; WAIDS: WAIS-III Digit Symbol Coding; WAIIN: WAIS-III Information; WAICA: WAIS-III Arithmetic; BENTFR: Benton Facial Recognition Test; Hint: Hinting Task; DFR_PHF: Degraded Facial Affect Recognition (DFAR) Percent happy faces; DFR_PFF: DFAR Percent fearful faces; DFR_PAF: DFAR Percent angry faces; DFR_PNF: DFAR Percent neutral faces; Significance levels: ***P <0 .001, **P <0 .01 and *P < 0.05.

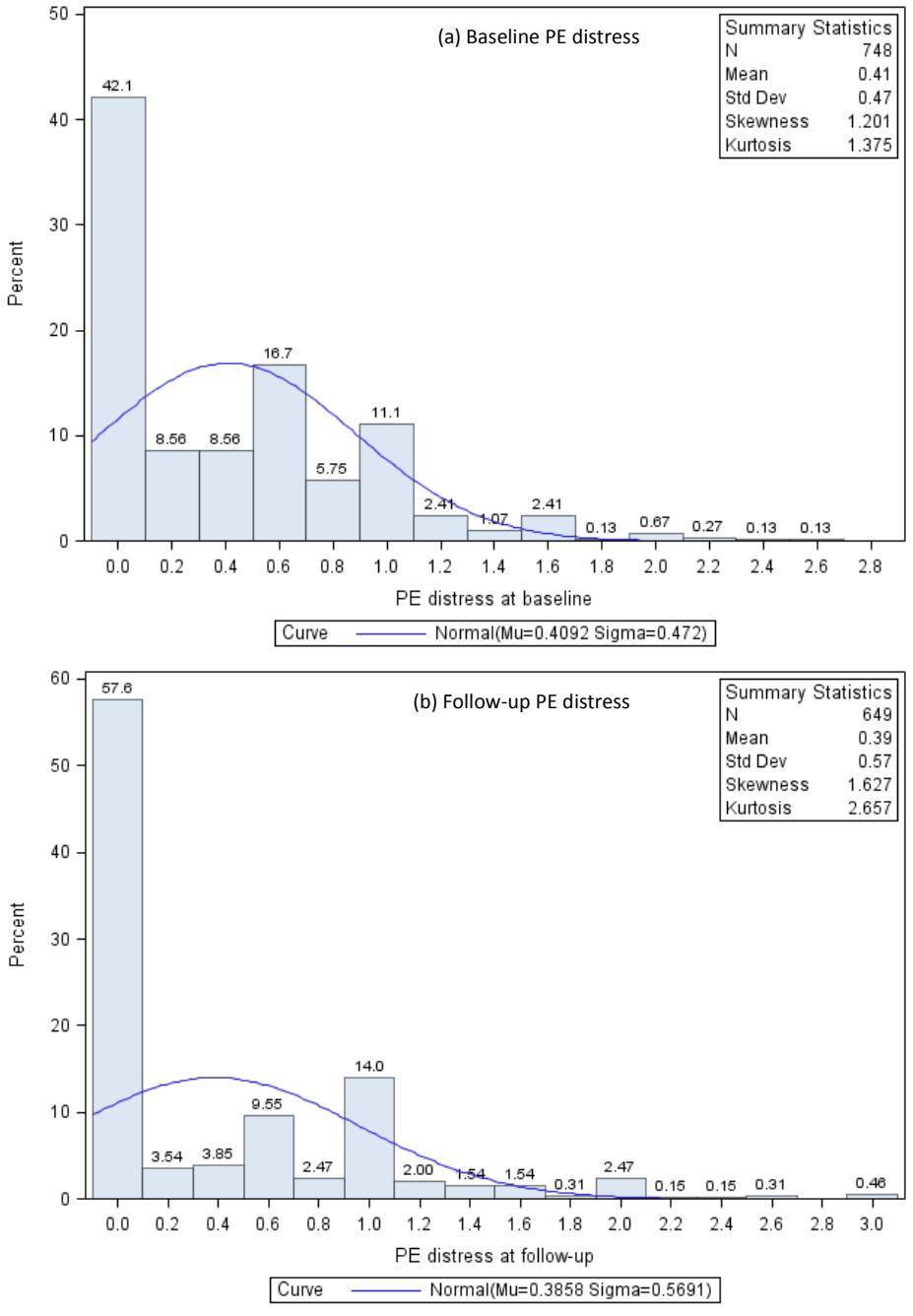
Supplementary Table S5: Pooled parameter estimates and model comparison for the frequency of positive psychotic experiences*

Parameter	Uncorrelated Model					Correlated Model**				
	Estimate	S.E.	95% C.I		P-value	Estimate	S.E.	95% C.I		P-value
			Lower	Upper				Lower	Upper	
Occurrence (Logistic)										
Intercept	0.946	1.377	-1.757	3.649	0.4923	0.877	1.245	-1.564	3.318	0.4813
Age	-0.013	0.023	-0.058	0.032	0.5711	-0.007	0.024	-0.055	0.041	0.7712
Sex	0.007	0.211	-0.407	0.420	0.9753	0.082	0.210	-0.333	0.497	0.6976
Education	-0.068	0.057	-0.180	0.045	0.2396	-0.083	0.056	-0.192	0.026	0.1339
Cannabis Use	-0.621	0.274	-1.158	-0.084	0.0234	-0.570	0.275	-1.114	-0.025	0.0404
Benton facial[a]	0.103	0.046	0.013	0.193	0.0245	0.091	0.041	0.011	0.170	0.0253
Time (3 years)	0.076	0.846	-1.590	1.741	0.9288	0.168	0.912	-1.713	2.048	0.8558
Immediate Recall[b]	0.047	0.027	-0.006	0.100	0.0809	0.054	0.025	0.004	0.103	0.0331
Immediate Recall*Time (3 years)	-0.068	0.030	-0.128	-0.009	0.0248	-0.077	0.032	-0.140	-0.013	0.0186
Var(random intercept): σ_1^2	2.791	0.685	1.447	4.135	<.0001	3.120	0.814	1.505	4.735	0.0002
Intensity (Lognormal)										
Intercept	-0.800	0.466	-1.718	0.117	0.0869	-0.741	0.430	-1.591	0.109	0.0871
Age	0.000	0.006	-0.012	0.012	0.9457	-0.002	0.006	-0.014	0.010	0.7648
Sex	0.028	0.053	-0.076	0.132	0.598	0.021	0.053	-0.083	0.126	0.6887
Education	-0.040	0.015	-0.071	-0.010	0.009	-0.043	0.018	-0.080	-0.006	0.0242
Cannabis Use	-0.108	0.065	-0.236	0.021	0.1007	-0.125	0.067	-0.257	0.007	0.0639
Sensitivity Index[c]	-0.006	0.003	-0.011	0.000	0.043	-0.006	0.003	-0.013	0.000	0.0444
Percent neutral faces[d]	-0.004	0.002	-0.007	-0.001	0.0238	-0.004	0.002	-0.007	-0.001	0.0177
Time (3 years)	0.543	0.427	-0.298	1.384	0.2051	0.582	0.438	-0.319	1.483	0.1957
Hinting Task	0.016	0.018	-0.020	0.051	0.3803	0.018	0.019	-0.020	0.056	0.3501
Hinting Task*Time (3 years)	-0.052	0.023	-0.096	-0.007	0.0233	-0.057	0.023	-0.104	-0.010	0.02
Var(res): σ_e^2	0.346	0.028	0.290	0.402	<.0001	0.325	0.022	0.280	0.369	<.0001
Var(random intercept): σ_2^2	0.256	0.031	0.195	0.317	<.0001	0.324	0.032	0.261	0.386	<.0001
Covariance	...	...	...	...	...	0.945	0.132	0.683	1.208	<.0001
Correlation (ρ)	...	...	...	...	...	0.941				
Model comparison (Fit Statistics)										
Criterion (Pooled)	Value					Value		Difference in -2LL		P-value
AIC	-275.30					-367.382				
-2LL	-319.30					-413.382		94.08		<.0001

*The model adjusted for age, gender, education, and cannabis used in past 12 months; [a]Benton Facial: Benton Facial Recognition Test; [b]Immediate Recall: WLT Immediate Recall; [c]Sensitivity Index: Continuous Performance Test Sensitivity Index; [d]Percent neutral faces: Degraded Facial Affect Recognition Task percent neutral faces; AIC: Akaike Information Criterion; -2LL: -2Log Likelihood. **Results of the correlated model are based on only four imputed datasets.



Supplementary Figure S1: Distribution of psychotic experiences (PE) positive frequency for siblings at (a) baseline and (b) 3-year follow-up



Supplementary Figure S2: Distribution of psychotic experiences (PE) distress for siblings at (a) baseline and (b) 3-year follow-up

CHAPTER 6

Familial liability to psychosis is a risk factor for multimorbidity in people with psychotic disorders and their unaffected siblings

Md. Atiqul Islam
M. Feroz Hayat Khan
Piotr J. Quee
Harold Snieder
Edwin R. van den Heuvel
Richard Bruggeman
Behrooz Z. Alizadeh
GROUP Investigators

Article in Press (European Psychiatry)

Abstract

Background: Multimorbidity may impose an overwhelming burden on patients with psychosis and is affected by gender and age. Our aim is to study the independent role of familial liability to psychosis as a risk factor for multimorbidity

Methods: We performed the study within the framework of the Genetic Risk and Outcome of Psychosis (GROUP) project. Overall, we compared 1024 psychotic patients, 994 unaffected siblings and 566 controls on the prevalence of 125 lifetime diseases, and 19 self-reported somatic complaints. Multimorbidity was defined as the presence of two or more complaints/diseases in the same individual. Generalized linear mixed model (GLMM) were used to investigate the effects of gender, age (adolescent, young, older) and familial liability (patients, siblings, controls) and their interactions on multimorbidity.

Results: Familial liability had a significant effect on multimorbidity of either complaints or diseases. Patients had a higher prevalence of multimorbidity of complaints compared to siblings (OR 2.20, 95% CI 1.79-2.69, P<0.001) and to controls (3.05, 2.35-3.96, P<0.001). In physical health multimorbidity, patients (OR 1.36, 95% CI 1.05–1.75, P=0.018), but not siblings, had significantly higher prevalence than controls. Similar finding were observed for multimorbidity of lifetime diseases including psychiatric diseases. Significant results were observed for complaints and disease multimorbidity across gender and age groups.

Conclusion: Multimorbidity is a common burden, significantly more prevalent in patients and their unaffected siblings. Familial liability to psychosis showed an independent effect on multimorbidity; gender and age are also important factors determining multimorbidity.

Keywords: familial liability; multimorbidity; physical health; psychosis; schizophrenia

1. Introduction

Schizophrenia spectrum disorder is a complex, multifaceted disorder with a 10-20 year shorter life expectancy (Laursen, 2011; Korver et al., 2012; Walker et al., 2015; Lawrence et al., 2013). A recent meta-analysis (Walker et al., 2015) on mortality in schizophrenia indicates that comorbid health-conditions may contribute up to ~67 percent of the triple excess of premature mortality in schizophrenia and related psychotic disorders (De Hert et al., 2011). Indeed, up to 54 percent of patients with schizophrenia have metabolic syndrome (Bruins et al., 2016; Vancampfort et al., 2015) and a 2-3 fold higher risk of diabetes mellitus (Bushe and Holt, 2004; van Winkel et al., 2006; Vancampfort et al., 2016) and cardiovascular diseases (Bresee et al., 2010; Hennekens et al., 2005; Hoang et al., 2013; Vancampfort et al., 2015) compared to general population. Precisely, type 2 diabetes mellitus is higher (2.9 percent) in patients with severe mental illness who prescribed antipsychotics (Vancampfort et al., 2016). Also, an extensive study indicates that schizophrenia patients have a broad range of multiple physical-health conditions but are less likely to have cardiovascular diseases than people without schizophrenia (Smith et al., 2013). Other diseases have also been identified in psychotic patients, including cancers, chronic obstructive pulmonary disease (COPD), tuberculosis, hepatitis C virus and osteoporosis (De Hert et al., 2011; Iacovides and Siamouli, 2008). It has been established that physical comorbidity in people with severe mental illness leads to a lower quality of life than for people without mental illness (Barnett et al., 2012; Payne et al., 2013; Qin et al., 2014; Reilly et al., 2015). However, little is known about the risk factors for the co-occurrence of physical illness, unexplained physical symptoms, and psychiatric diseases.

Multimorbidity is defined as any co-occurrence of two or more medical conditions within the same person (Batstra et al., 2002; Feinstein, 1970; Tomasdottir et al., 2013). It affects more than two thirds of patients in general practice and half of the elderly population (Schafer et al., 2012). So far most studies (Bresee et al., 2010; De Hert et al., 2011; Hennekens et al., 2005; Taylor et al., 2010; van Winkel et al., 2006) have focused on lifetime diseases but not on medical complaints, even though these medical complaints in themselves may also lead patients to seek medical attention. Multiple factors play a role in multimorbidity: *e.g.* the side effects of psychotropic drugs, associations with lifestyle risk factors, particularly smoking and substance abuse, and inequitable access to preventative health care and medical treatment (De Hert et al., 2011; Lawrence and Kisely, 2010). Above all the factors gender and age are considered the main determinants of multimorbidity for many diseases (Agborsangaya et al., 2012; Britt et al., 2008; Fuchs et al., 2012; Rizza et al., 2012; van den Akker et al., 1998; van Oostrom et al., 2012; Schafer et al., 2012). However, with regard to schizophrenia, the role of gender and age in predisposing to multimorbidity in psychosis and in healthy population is unclear.

Notably, psychosis is more prevalent in men and has a typical onset in younger people between 20 and 40 years (McGrath et al., 2004; Batki et al., 2009; Taylor et al., 2010; Korver et al., 2012; Manuel et al., 2013; Smith et al., 2013). Men are more likely than women to have lifetime risk until age 45-50 years. After that age on (50-54 years), women reach almost the same as men lifetime risk. Consequently, lower age groups are certain to lead to a male predominance in the risk ratios (Häfner, 2003; Abel et al., 2010). Physical multimorbidity in psychosis or healthy controls, on the other hand, more often occurs in women (Schafer et al., 2012; Smith et al., 2013; Agur et al., 2016;

Stubbs et al., 2016) and is more typically seen in older people than in younger adults (Marengoni et al., 2011; Smith et al., 2013; Agur et al., 2016; Stubbs et al., 2016). If therefore, multimorbidity in psychosis has a different man/woman ratio, and that multimorbidity in psychosis emerges at a younger age, we can argue that genetic factors, family links, environmental factors (e.g. high rate of smoking and alcohol use in both patients and siblings) may play a greater role in liability to psychosis. This would suggest that family liability to psychosis is another important determinant for multimorbidity.

To study the effect of familial liability, we will evaluate a cohort of young psychotic patients and their unaffected siblings. Unaffected siblings of people with psychosis report increased numbers of subclinical psychotic symptoms (Chen et al., 2009; Genetic Risk and Outcome in Psychosis (GROUP) Investigators, 2011), have a 10-fold increased risk of developing schizophrenia (Gejman et al., 2011), and share ~50% of their genes with their ill relative, including a number of the schizophrenia risk genes (Gottesman and Gould, 2003). A large multinational study demonstrates that physical health multimorbidity is increased across the control, with subclinical psychosis and psychosis in 48 low- and middle-income countries (Stubbs et al., 2016). The study of siblings allows us to investigate the prevalence of multimorbidity in persons who to some extent share familial liability for psychosis, but by whom multimorbidity is related neither to the course of schizophrenia nor the use of antipsychotics. We will investigate the effects of gender, age, and familial liability on the prevalence of multimorbidity (of complaints, physical health and lifetime diseases including psychiatric diseases) and their possible two-way (*gender* x *age*; *age* x *familial liability*; *gender* x *familial liability*) and three-way (*gender* x *age* x *familial liability*) interactions.

The present study will examine (*i*) whether the liability to psychosis overrules the effect of gender and age on multimorbidity, whereby younger men with psychosis have an increased rate of multimorbidity, and (*ii*) whether the same effects hold for their unaffected siblings.

2. Methods

2.1. Design and setting of the study

The study was conducted within the framework of the Genetic Risk and Outcome of Psychosis (GROUP) project (Data release version 3.02), a large prospective cohort study in the Netherlands and Belgium. The GROUP project has been described in detail elsewhere (Korver et al., 2012). From a total of 3684 participants at baseline, we included 2584 participants (1024 patients, 994 siblings, and 566 controls) in the present study. Persons identified as potentially eligible were given detailed explanation of the study procedures and were asked for informed consent for detailed assessment and for contacting their first-degree family members (brothers, sisters, parents). Controls were selected through a system of random mailings to addresses in the catchment areas of the cases (Korver et al., 2012). The GROUP study protocol was approved centrally by the Ethical Review Board of the University Medical Centre Utrecht and subsequently by local review boards of each participating institute (Korver et al., 2012).

The selection of the sample is depicted in Supplementary Figure S1 and explained in Supplementary Methods. Participants were divided into three subgroups based on inheritance of liability to psychosis: 1) patients with a highest familial liability, 2) their corresponding unaffected

singleton of multiple siblings with an increased familial liability and 3) unrelated controls with no family history of psychosis as those with the lowest familial liability of psychosis.

2.2. Population demography

Socio-demographic characteristics used in the analysis were age, gender, ethnicity, educational attainment (adapted from Verhage (Verhage, 1964)), intelligence quotient (IQ) as estimated by The Wechsler Adult Intelligence Scale, third edition (WAIS-III) (Wechsler et al., 2008), marital and residential status, gross monthly income at the time of recruitment, smoking and alcohol consumption.

2.3. Characterization of complaints and lifetime diseases

Generally, a complaint is a symptom of which a person is aware or which causes discomfort. It is usually described from a patient's perspective and is often his/her key reason for seeking medical attention. On the other hand, lifetime disease is an illness or sickness characterized by specific signs and symptoms which people have developed and had diagnosed during their lifetime (Vos et al., 2013). Data on medical conditions, including complaints and lifetime diseases, were derived from two main resources: a medical questionnaire and pharmacy records. A Medical Questionnaire was designed for interviewing participants to identify their somatic comorbidities. It was noted that self-reported diseases were comparable with objectively diagnosed diseases in population-based studies (Lampe et al., 1999). We divided the questionnaire into two sections, including a 'checklist' of multiple choice questions and a 'narrative section,' consisting of open questions. The thirty-eight medical conditions from the checklist section were divided into 19 complaints and 19 lifetime diseases according to the 10th revision of the International Classification of Diseases (ICD-10). Another 102 lifetime diseases were retrieved from the narrative section; these were matched with the checklist of lifetime diseases to avoid duplication. Similarly, complaints were removed from the checklist if participants had received a medical diagnosis (e.g. lifetime diseases). In total, 121 physical health morbidity (excluding mental and behavioural disorders) were retrieved from the database. Note that at baseline, five groups of psychiatric disorders (psychotic disorder: schizophrenia, schizoaffective and delusional; mood; anxiety and/or eating; substance, and others) were formed regarding patients, siblings and controls, respectively. All patients were diagnosed with non-affective psychosis. The details has been depicted by a flow diagram in the Supplementary Figure S1. For simplicity of presentation, the observed 125 lifetime diseases including four mental and behavioural disorders were categorized into 16 disease domains based on the bodily organ system to which each disease belonged (Supplementary Table S1-S3).

2.4. Outcomes measure

For each subject, we counted the number of complaints and lifetime diseases (including and excluding mental and behavioural disorders) independently. We defined multimorbidity of complaints/diseases as the presence of two or more complaints/lifetime diseases in the same individual. We treated multimorbidity of complaints, physical morbidity and lifetime diseases as three separate outcomes throughout the study and as binary outcomes (i.e. present/absent). We did

not count psychotic disorder (e.g. schizophrenia, schizoaffective and delusional) in patients, as this was already their indexed disease. In summary, we collected 19 complaints and 125 lifetime diseases to measure multimorbidity regarding complaints, physical health multimorbidity and lifetime diseases separately (Supplementary Figure S1, Table S1 and Table S2).

2.5. Data analysis

We compared psychotic patients, their unaffected siblings and controls pairwise, for socio-demographic characteristics. We conducted independent group comparisons using Chi-square, T-test or Mann-Whitney U test, depending on the nature of variables; we compared correlated groups (e.g. patients and siblings) using the McNemar test or the Generalized Estimating Equations (GEEs) model. Moreover, we estimated the prevalence of multimorbidity of complaints and lifetime diseases in terms of percentages across familial liability groups by gender and age groups: adolescents ≤ 20 years, young adults 21-40 years, and older adults >40 years. Because age range is *a priori* factor involving more risk of multimorbidity, we categorized age into three meaningful groups (Batki et al., 2009; Marengoni et al., 2011; Manuel et al., 2013).

Since patients and siblings belonged to the same family, individuals within families were expected to be more homogenous than between families. To investigate the effects of gender, age, familial liability and their interactions on the prevalence of multimorbidity of complaints, physical health and lifetime diseases, we conducted a Generalized Linear Mixed Model (GLMM) (Molenberghs and Verbeke, 2006), taking the family into account as a random effect. We used an adaptive Gaussian quadrature with 20 quadrature points to estimate the parameters and their associate standard errors. If an interaction effect was not statistically significant, we reported only the main effect. In the model building process, we incorporated main (gender, age group and familial liability) and all possible two- and three-way interaction effects into the full model. We performed a backward elimination procedure and compared the models, using the corrected Akaike's Information Criterion (AICc) and the Bayesian information criterion (BIC). The smallest values of AICc and BIC concluded the final model. We used Type-III overall tests of fixed effect P-values to conclude the marginal effects of complaints and lifetime diseases on multimorbidity separately. We also conducted pairwise comparisons on main effects if they were significant. We also calculated the intraclass correlation coefficient (ICC) to measure the homogeneity of subjects within a family. We conducted all analyses using Statistical Analysis System (SAS Institute Inc., Cary, NC) version 9.4 with a two-tailed test at 5% level of significance.

3. Results

3.1. Descriptive of study population

Table 1 describes and compares the basic socio-demographics, risk factors and other subclinical characteristics among patients, siblings and controls. Overall, patients (age: mean 27.8± SD 8.2) and siblings (27.8±8.2) were significantly (P<0.001) younger than controls (30.5±10.6). In patients, the proportion of women (24.5%) was significantly lower (P<0.001) compared to siblings (54.3%) and controls (55.5%). Patients and siblings had significantly (P<0.001) less education, lower IQ's, and

higher use of alcohol and nicotine compared to controls. Moreover, 86.5% patients used antipsychotics while other two groups e.g. siblings and controls did not use any antipsychotic.

The prevalence of multimorbidity of complaints, physical and lifetime diseases across the familial liability subgroups (i.e. patients, siblings and controls) respectively, are depicted in the Figure 1-3. The occurrence of four or more complaints was 24.9%, 12.6% and 9.0% for patients, siblings and controls respectively. Multimorbidity of complaints was prevalent in 51.5% of the patients, 35.7% of the siblings and 31.3% of the controls (Figure 1). Physical health comorbidity was observed in 46.1% of the patients, 46.2% of the siblings and 45.6% of the controls (Figure 2). Multimorbidity of lifetime diseases was found in 47.2% of the patients, 47.0% of the siblings and 46.1% of the controls (Figure 3).

The observed prevalence of multimorbidity of complaints, physical health and lifetime diseases across familial liability subgroups by gender and age groups respectively are presented in Supplementary Results and are demonstrated in the Supplementary Figures S2-S4 and Tables S1-S4.

Table 1: Baseline socio-demographic characteristics for patients, siblings and controls[a].

Characteristics	Familial liability group		
	Control (N=566)	Sibling (N=994)	Patient (N=1024)
Age in years			
Overall, mean±sd (Range)	30.5±10.6 (15-56)	27.8±8.3 (14-60)^{3†}	27.8±8.2 (15-68)^{3†}
Age in category (years)			
Adolescents, N (mean±sd)	130 (18.2±1.2)	204 (18.1±1.6)	163 (18.5±1.4)$^{2\dagger 2*}$
Young adults, N (mean±sd)	300 (29.0±6.0)	703 (28.3±5.2)	778 (27.7±5.2)$^{2\dagger 2*}$
Older adults, N (mean±sd)	136 (45.6±3.3)	87 (45.8±4.3)	83 (46.7±5.9)
Gender, % Women	55.5	54.3	24.5$^{3\dagger 3*}$
Ethnicity, % Caucasian	92.1	83.7^{3†}	79.0$^{3\dagger 3*}$
Education[1], mean±sd	5.4±1.8	5.1± 2.1^{3†}	4.1±2.1$^{3\dagger 3*}$
IQ, estimated[2], mean±sd	109.6±14.7	102.8±15.4^{3†}	95.3 ±16.1$^{3\dagger 3*}$
Married/Living together, %	40.7	39.9	9.5$^{3\dagger 3*}$
Living with partner/family, %	46.5	46.0	10.7$^{3\dagger 3*}$
Income minimal, % (€<1264.80)	35.3	28.9†	63.0$^{3\dagger 3*}$
Nicotine use, %	27.1	41.5^{3†}	61.9^{3†}
Alcohol use, %	62.7	73.7^{3†}	71.2$^{3\dagger 3*}$
Current use of Antipsychotics, %	-	-	86.5

[a]Table presents Mean±SD or numbers (in %); Adolescents ≤ 20 years, young adults = 21-40 years, older adults >40 years; [1]Education (Verhage): range 0 (no education), 3-5 (school diploma) to 8 (university degree); [2]IQ: Wechsler Adult Intelligence Scale-III (WAIS-III).

†Patient or sibling compared to control (using Chi-square and/or Mann-Whitney test): †P<0.05, ^{2†}P<0.01 and ^{3†}P<0.001. *Patient compared to sibling (using Generalized Estimating Equations): ^{2*}P<0.01 and ^{3*}P<0.001.

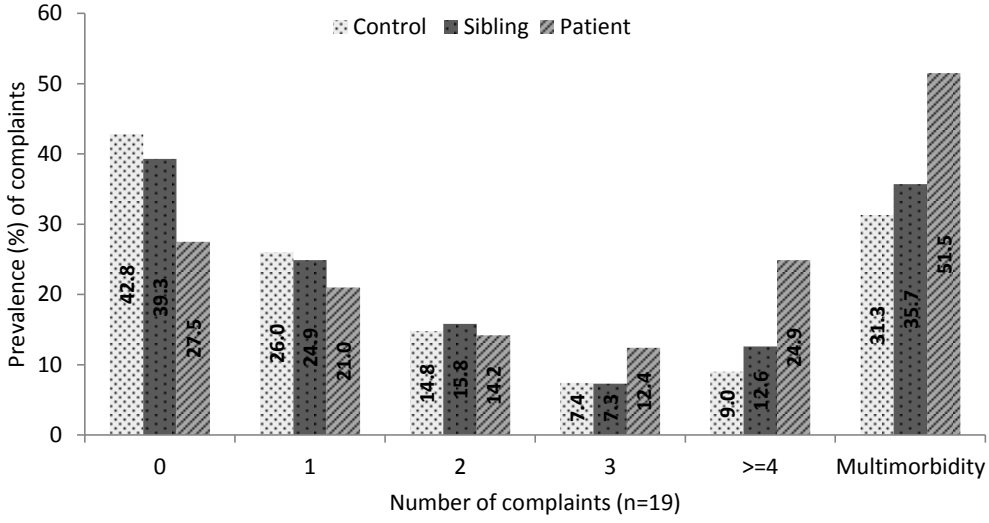


Figure 1: Prevalence of occurrence of complaints for different groups of subjects.

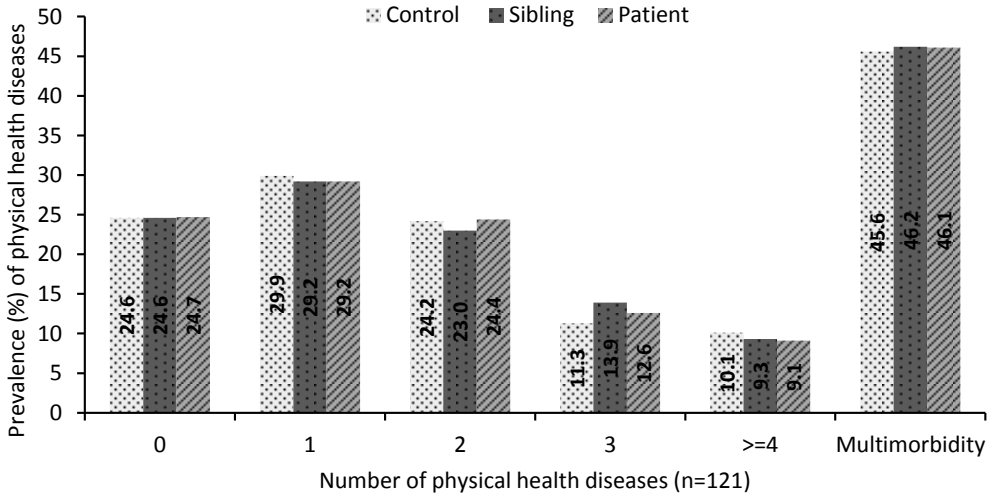


Figure 2: Prevalence of occurrence of physical health diseases for different groups of subjects.

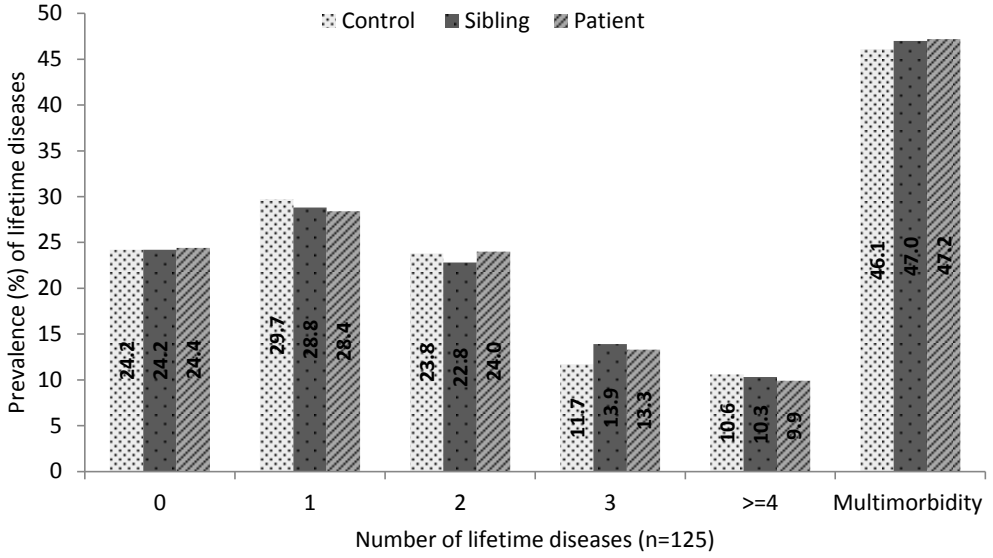


Figure 3: Prevalence of occurrence of lifetime diseases for different groups of subjects.

3.2. Effects of gender, age groups and familial liability on the prevalence of multimorbidity

We found no significant interaction effects of gender, age groups and familial liability groups on the prevalence of multimorbidity, either of complaints or physical health multimorbidity or lifetime diseases. The main effects from type III overall tests indicated that gender (P=0.003), age group (P<0.001) and familial liability (P<0.001) were significantly related to the prevalence of multimorbidity of complaints. Table 2 presents the pairwise comparisons, odds ratios, their 95% confidence interval and ICC for multimorbidity. In model-1, both being a patient with the highest liability of psychosis (OR 3.05, 95% CI 2.35-3.96, P<0.001) and being a sibling with increased familial liability of psychosis (OR 1.39, 95% CI 1.08–1.78, P=0.01) had a significant effect on having more complaints, as compared to those with the lowest familial liability for psychosis (i.e. controls). Also, patients had significantly, 2.20 (95% CI 1.79-2.69, P<0.001), times more multimorbid complaints than their corresponding siblings. Older adults had 2.24 times (95% CI 1.60–3.13, P<0.001) higher prevalence of multimorbidity of complaints compared to adolescents. Older adults had more multimorbid complaints than young adults (OR 1.86, 95% CI 1.39-2.47, P<0.001). The odds of women having multimorbidity of complaints were 1.33 (95% CI 1.11–1.61, P=0.003) times higher than for men (Table 2, model-1).

Similarly, in model-2, psychotic patients had 1.36 (95% CI 1.05–1.75, P=0.018) times higher odds for physical health multimorbidity compared to controls. Young adults and older adults were significantly associated with 1.55 (95% CI 1.23–1.97, P<0.001) and 3.83 (95% CI 2.69–5.46, P<0.001) times higher for physical multimorbidity than adolescents. Women had significantly higher (OR 1.70, 95% CI 1.41-2.06, P<0.001) physical health multimorbidity than for men (Table 2, model-2).

Additionally, in model-3, type III overall tests of fixed effects showed that gender (P<0.001), age group (P<0.001) and familial liability (P=0.038) were significantly associated with the prevalence of multimorbidity of lifetime diseases. Patients had 1.39 (95% CI 1.08–1.79, P=0.011) times higher prevalence of lifetime diseases compared to controls. Young adults (OR 1.59, 95% CI 1.25–2.02, P<0.001) and older adults (OR 4.06, 95% CI 2.84–5.80, P<0.001) had significantly more disease multimorbidity than adolescents. Older adults also had more multimorbid diseases than young adults (OR 2.56, 95% CI 1.88-3.47, P<0.001). Likewise, the odds of women having disease multimorbidity were 1.68 (95% CI 1.39–2.03, P<0.001) times higher than for men (Table 2, model-3).

The intraclass correlation coefficients (ICC) between subjects within a family were 0.11 (95% CI 0.03-0.19) for multimorbidity of complaints, 0.15 (95% CI 0.07-0.23) for physical health multimorbidity and 0.16 (95% CI 0.08-0.24) for multimorbidity of lifetime diseases, meaning that proportion of the total variance (11%, 15% and 16%) in multimorbidity of complaints, physical and lifetime diseases, respectively, that were accounted for by the family effect (Table 2).

Table 2: Pairwise comparisons, odds ratios, their 95% confidence interval and ICC for multimorbidity[a].

Factors	Odds Ratio (95% C.I)	P-value
Model-1: Multimorbidity of complaints		
Patient vs. Control	3.05 (2.35-3.96)	<0.001
Sibling vs. Control	1.39 (1.08-1.78)	0.010
Patient vs. Sibling	2.20 (1.79-2.69)	<0.001
Young adult vs. adolescent	1.21 (0.96-1.52)	0.114
Older adult vs. adolescent	2.24 (1.60-3.13)	<0.001
Older adult vs. young adult	1.86 (1.39-2.47)	<0.001
Women vs. Men	1.33 (1.11-1.61)	0.003
Model-2: Physical health multimorbidity (excluding psychiatric diseases)		
Patient vs. Control	1.36 (1.05-1.75)	0.018
Sibling vs. Control	1.18 (0.92-1.52)	0.188
Patient vs. Sibling	1.15 (0.94-1.41)	0.178
Young adult vs. adolescent	1.55 (1.23-1.97)	<0.001
Older adult vs. adolescent	3.83 (2.69-5.46)	<0.001
Older adult vs. young adult	2.46 (1.82-3.34)	<0.001
Women vs. Men	1.70 (1.41-2.06)	<0.001
Model-3: Multimorbidity of lifetime disease (including psychiatric diseases)		
Patient vs. Control	1.39 (1.08-1.79)	0.011
Sibling vs. Control	1.21 (0.94-1.55)	0.143
Patient vs. Sibling	1.15 (0.94-1.41)	0.165
Young adult vs. adolescent	1.59 (1.25-2.02)	<0.001
Older adult vs. adolescent	4.06 (2.84-5.80)	<0.001
Older adult vs. young adult	2.56 (1.88-3.47)	<0.001
Women vs. Men	1.68 (1.39-2.03)	<0.001
Variance:	Estimate	ICC[b] (95% C.I)
Model-1: Intercept (Family effect)	0.40	0.11 (0.03-0.19)
Model-2: Intercept (Family effect)	0.59	0.15 (0.07-0.23)
Model-3: Intercept (Family effect)	0.60	0.16 (0.08-0.24)

[a]Table presents the GLMM pairwise comparison results; Adolescents ≤ 20 years, young adults = 21-40 years, older adults>40 years; [b]ICC: Intraclass correlation coefficient = [Variance of random intercept/(Variance of random intercept+π^2/3)]. Model-1: AICc=3384.07, BIC=3421.12; Model-2: AICc= 3459.76, BIC= 3496.81.; Model-3: AICc=3462.42, BIC=3499.47.

4. Discussion

This study described the prevalence of multimorbidity in psychotic patients, their unaffected siblings and controls, and investigated the role of familial liability to psychosis as a determinant of multimorbidity, next to gender and age.

In all three groups, women showed a higher prevalence of multimorbidity for both complaints and diseases with/without psychiatric diseases, and the prevalence of multimorbidity increased with age. Also, the prevalence of multimorbidity was increased by the degree of familial liability for psychosis. Studying the two- and three-way interactions, this study demonstrates familial liability to psychosis to be a consistent determinant of the risk of multimorbidity across gender and age groups. For medical complaints, we found the pain of joints or muscles, allergy, problematic bowel movement, dizziness and palpitation to be the most prevalent medical complaints of all three groups. These complaints also prevailed in general population studies (den Boeft et al., 2016). Note that pain prevalence of psychotic patients in the current study did not differ significantly either from controls or siblings, which was in line with the recent meta-analysis studied by Stubbs et al (Stubbs et al., 2014). Since medical complaints were based on self-reported data, lack of pain sensitivity measure yielded the same conclusion, which might be under-estimate. There might be other factors e.g. impaired prefrontal, medial temporal functioning (Keshavan et al., 2008), use of antipsychotics (Seidel et al., 2013; Stubbs et al., 2015) and alterations in the cortical dopamine system (Jarcho et al., 2012) which reduced pain sensitivity in people with schizophrenia. However, a meta-analysis of clinical pain induction studies demonstrated that higher psychiatric symptoms moderated increased pain threshold, and younger patient age moderated increased pain tolerance (Stubbs et al., 2015).

Of all the comorbid diseases, we found concussion, eczema, migraine, tonsillitis, congenital defects, mood disorders and COPD across the familial liability groups to be the most prevalent lifetime diseases. Other diseases had an occurrence of less than 5% (Supplementary Table S1 and S4). The current study found higher levels of migraine in siblings and controls but not in patients. This might be due to under-reporting clinical symptoms and diseases in patients which reduced pain sensitivity as patients used antipsychotics, higher levels of nicotine and alcohol.

Different operational definitions of multimorbidity have been used in the literature (Willadsen et al., 2016). The majority of studies have defined multimorbidity as two or more, whereas others counted three or more, concurrent diseases (Fortin et al., 2005; Jacobi et al., 2004; Fuchs et al., 2012). A medical complaint is in itself a symptom or a set of symptoms, not a disease. The present study extracted 19 such complaints or symptoms independently and had formed a separate definition of multimorbidity of complaints, leading to another dimension for studying multimorbidity and finding their determinants. This approach is in line with the suggestions of Willadsen et al., based on the data from their recent meta-analysis, including severity and symptoms, making the existing definitions more suitable for epidemiologists than for clinicians (Willadsen et al., 2016). Along with this meta-analysis, the separation of multimorbidity into complaints, physical health and lifetime diseases has made the present study more comprehensive and clinically relevant. Examining comorbidity through the mechanism of multimorbidity allows for evaluation of the overall health of a familial liability group and avoids the complexity of comparing many specific diseases, particularly those with a low prevalence. The current study was the first to make a more clear

distinction between complaints and diseases, thus providing a novel comparison of physical health between patients with non-affective psychosis and their unaffected siblings.

In our study, healthy controls reported overall prevalence of 31.3 percent multimorbidity for complaints, 45.6 percent for physical health, and 46.1 percent for lifetime diseases. A recent systematic review demonstrated a prevalence rate of 13.1 to 71.8 percent for the general population across studies (Fortin et al., 2012). For example, in Switzerland, the prevalence of multimorbidity ranged from 47 to 97 percent in medical inpatient records (Schneider et al., 2012), comparable to 36 to 43 percent of the older population in a German national health survey (Fuchs et al., 2012). Other studies showed a lower prevalence of multimorbidity, such as the 17 percent frequency of all ages based on an Australian biomedical cohort study (Taylor et al., 2010). In the Netherlands, Van den Akker et al. (1998) reported a 29 percent overall multimorbidity of diseases in Registration Network Family Practice data (van den Akker et al., 1998). Although comparisons with other cohorts were hampered by a lack of a standard definition of multimorbidity, the percentages of complaints and diseases in our healthy controls were within the range of studies in the general population, adding to the validation of our sample. When looking at gender, several population based studies (Agborsangaya et al., 2012; Britt et al., 2008; Marengoni et al., 2011; O'Kelly et al., 2011; Rizza et al., 2012; van Oostrom et al., 2012) reported an increased proportion of multimorbidity in women compared to men in general. Additionally, gender difference was also observed in three other studies (Smith et al., 2013; Crump et al., 2013; Stubbs et al., 2016) where women with schizophrenia were more likely to have physical multimorbidity than men. Interestingly, women showed a tendency to have a longer life expectancy, but also a lower quality of health compared to men (Foguet-Boreu et al., 2014; Ha et al., 2015). In line with these literatures, our study also showed multimorbidity to be more common in women than in men. In our study, the gender difference might be explained by the fact that women were older on average (29.37±9.6) than men (27.59±8.22) and women more often consult with their general practitioner (Smith et al., 2013; Crump et al., 2013). We also observed this pattern across familial liability classes, meaning that women more often had multimorbidity of diseases and complaints than did men. However, within both genders, patients more often had multimorbidity than their siblings and the latter more often than controls. Thus, gender was confirmed as a major determinant of multimorbidity. At the same time, these data suggested that the effect of familial liability was independent of gender.

Like previous studies, our study showed that age was associated with multimorbidity of complaints as well as of physical and lifetime diseases (Agborsangaya et al., 2012; Schafer et al., 2012; Stubbs et al., 2016). When studying older control groups, we observed 42.3 to 69 percent multimorbidity, which differed from previous studies reporting 32 to 36 percent for the 40-59 year age group (Taylor et al., 2010; Marengoni et al., 2011). We also found that the prevalence of lifetime disease multimorbidity increased with increasing age. In older patients, we found a 68.1 to 72.2 percent multimorbidity of lifetime diseases for both genders (mainly within 40-50 years), which was higher than age-matched unaffected siblings and controls; further unaffected siblings showed trends similar to those of controls when compared to the findings from previous studies (Fortin et al., 2005; Marengoni et al., 2011; Taylor et al., 2010). Thus, we found that having schizophrenia showed an effect on multimorbidity independent of age, suggesting that the presence of schizophrenia may

trigger a primed susceptibility to multiple diseases at a younger stage of life. In physical health multimorbidity, patients but not siblings had more risk on having diseases than controls. This result was in line with a recent large study confirming that schizophrenia patients had multiple physical health diseases than healthy people (Smith et al., 2013). Also, prevalence of lifetime diseases including psychiatric diseases demonstrated a slightly increasing trend (controls<siblings<patients), but here only the difference between patients and controls reached statistical significance (Figure 3). Regarding the multimorbidity of complaints, patients had a significantly higher number of complaints than their siblings, who in turn had a significantly higher number of complaints than controls. Therefore, the estimated prevalence of multimorbidity was higher in patients with schizophrenia compared to siblings and controls, but it followed similar increasing trends with ageing, as reported in other studies (Agborsangaya et al., 2012; Britt et al., 2008; O'Kelly et al., 2011; Rizza et al., 2012; Schneider et al., 2012; van Oostrom et al., 2012). A large multi nation-wide study also showed that the prevalence of physical health multimorbidity was increased across the psychosis-spectrum e.g. 11.4 in controls, 21.8 percent in subclinical psychosis (more than one psychotic symptom in the past 12 months, but no lifetime diagnosis of psychosis) and 36.0 percent in lifetime diagnosis of psychosis. Although the prevalence was lower than the current study but the risks of subclinical psychosis and psychosis for multimorbidity were comparable with the present study (Stubbs et al., 2016).

Although this study was based on a large cohort of patients and their unaffected siblings, these data needed to be interpreted with some caution. This study focused on the main determinants of gender, age and new discoveries of familial liability, leaving out other reported multiple risk factors (Agborsangaya et al., 2012; De Hert et al., 2011) for multimorbidity. We showed that familial liability is one of the main determinants of multimorbidity which acts independently of the other major risk factors of age and gender. It is essential to perform additional genetic analysis to distinguish between the effects of genetic liability and intra-familial environmental susceptibility. Some other risk factors, *e.g.* psychotropic drugs, applied only to the patients; others such as nicotine, alcohol use, and IQ were described in the population characteristics. The present study was based mainly on self-reported diseases, clinical complaints, and medication history. For the selected population with psychotic disorders, there might be a reporting bias due to the diagnosis of the disease. Comparison to their siblings and matched healthy controls could lead to underrepresentation of the numbers, rather than an indication of the presence of other psychiatric symptoms like mood disorders, anxiety or eating disorders. A major methodological concern is how to deal with other psychiatric disorders. The inclusion of those disorders could be problematic; *e.g.* it is not uncommon for people with psychotic disorders to receive a diagnosis of mood disorder at some time in their lives because of the different presentation of symptoms at different times in the course of the disease. What presents may not always be a distinct disease but may have symptoms of the psychotic disorder. Mood disorder in patients was the most frequently co-occurring psychiatric disorder. However, only 5.5% of the patients received a mood disorder diagnosis, while depressive symptoms were more prevalent (62.0%) indicating that in our sample depressive symptoms are considered by and large to be an integrated part of the psychotic disorder. It must be noted that we did not include psychotic disorders in patients while estimating multimorbidity, as we wished to give the latter similar weight as in the other participants (siblings and controls). Another concern is that

we formed anxiety and eating disorders together as anxiety disorders co-occur with eating disorders. In fact, the existence of anxiety disorders can often lead to the development of an eating disorder i.e. anxiety precedes an eating disorder. Although anxiety and eating disorders are in different diagnosis groups with high comorbidity, both act similar clinical features (Hocaoglu, 2017).

The strength of this study was that it allowed for investigation of a large number of lifetime diseases and complaints in a large sample of psychotic patients, siblings, and controls. In fact, no previous study has dealt with a comprehensive list of diseases, either somatic or psychiatric, in people with psychotic disorders while counting multimorbidity in a cumulative way rather than focusing only on pairwise comorbidities (Nuyen et al., 2006; Oreski et al., 2012). The most informative feature of this study was the sibling model; whereas most studies emphasized multimorbidity either in healthy subjects or the disease population, this is the first study to consider multimorbidity of siblings of people with psychotic disorders.

4.1. Conclusions

Multimorbidity is a clinical burden not only for people with psychotic disorders but also for their family members. In psychosis, multimorbidity was strongly associated with women and was not limited to the elderly but also affected young individuals. This information may be vital to quantify the burden of multimorbidity on patients and, most importantly, on unaffected siblings. The risk of multimorbidity caused by familial liability for psychosis was consistent across gender and age group, meaning that familial liability for psychosis is in itself a strong independent determinant of multimorbidity. Future studies could investigate the role of heritability and genetic factors in multimorbidity to confirm our findings and gain more insight into causal factors.

References

Abel, K.M., Drake, R., Goldstein, J.M. (2010). Sex differences in schizophrenia. *International Review of Psychiatry* 22:417-428.

Agborsangaya, C.B., Lau, D., Lahtinen, M., Cooke, T., Johnson, J.A. (2012). Multimorbidity prevalence and patterns across socioeconomic determinants: a cross-sectional survey. *BMC public health* 12:201.

Agur, K., McLean, G., Hunt, K., Guthrie, B., Mercer, S.W. (2016). How Does Sex Influence Multimorbidity? Secondary Analysis of a Large Nationally Representative Dataset. *International Journal of Environmental Research and Public Health* 13:391.

Barnett, K., Mercer, S.W., Norbury, M., Watt, G., Wyke, S., Guthrie, B. (2012). Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *The Lancet* 380:37-43.

Batki, S.L., Meszaros, Z.S., Strutynski, K., Dimmock, J.A., Leontieva, L., Ploutz-Snyder, R., Canfield, K., Drayer, R.A. (2009). Medical comorbidity in patients with schizophrenia and alcohol dependence. *Schizophrenia research* 107:139-146.

Batstra, L., Bos, E.H., Neeleman, J. (2002). Quantifying psychiatric comorbidity--lessions from chronic disease epidemiology. *Social psychiatry and psychiatric epidemiology* 37:105-111.

Bresee, L.C., Majumdar, S.R., Patten, S.B., Johnson, J.A. (2010). Prevalence of cardiovascular risk factors and disease in people with schizophrenia: a population-based study. *Schizophrenia research* 117:75-82.

Britt, H.C., Harrison, C.M., Miller, G.C., Knox, S.A. (2008). Prevalence and patterns of multimorbidity in Australia. *The Medical journal of Australia* 189:72-77.

Bruins, J., Pijnenborg, M.G., Bartels-Velthuis, A.A., Visser, E., van den Heuvel, E.R., Bruggeman, R., Jorg, F. (2016). Cannabis use in people with severe mental illness: The association with physical and mental health--a cohort study. A Pharmacotherapy Monitoring and Outcome Survey study. *Journal of psychopharmacology (Oxford, England)* 30:354-362.

Bushe, C., Holt, R. (2004). Prevalence of diabetes and impaired glucose tolerance in patients with schizophrenia. *The British journal of psychiatry.Supplement* 47:S67-71.

Chen, L.S., Rice, T.K., Thompson, P.A., Barch, D.M., Csernansky, J.G. (2009). Familial aggregation of clinical and neurocognitive features in sibling pairs with and without schizophrenia. *Schizophrenia research* 111:159-166.

Crump, C., Winkleby, M.A., Sundquist, K., Sundquist, J. (2013). Comorbidities and mortality in persons with schizophrenia: a Swedish national cohort study. *The American Journal of Psychiatry* 170:324-333.

De Hert, M., Correll, C.U., Bobes, J., Cetkovich-Bakmas, M., Cohen, D., Asai, I., Detraux, J., Gautam, S., Moller, H.J., Ndetei, D.M., Newcomer, J.W., Uwakwe, R., Leucht, S. (2011). Physical illness in patients with severe mental disorders. I. Prevalence, impact of medications and disparities in health care. *World psychiatry : official journal of the World Psychiatric Association (WPA)* 10:52-77.

De Hert, M., Dekker, J.M., Wood, D., Kahl, K.G., Holt, R.I., Moller, H.J. (2009). Cardiovascular disease and diabetes in people with severe mental illness position statement from the European Psychiatric Association (EPA), supported by the European Association for the Study of Diabetes (EASD) and the European Society of Cardiology (ESC). *European psychiatry : the journal of the Association of European Psychiatrists* 24:412-424.

den Boeft, M., Twisk, J.W.R., Terluin, B., Penninx, B.W.J.H., van Marwijk, Harm W. J., Numans, M.E., van, d.W., van, d.H. (2016). The association between medically unexplained physical symptoms and health care use over two years and the influence of depressive and anxiety disorders and personality traits: a longitudinal study. *BMC Health Services Research* 16:1-9.

Feinstein, A.R. (1970). The pre-therapeutic classification of co-morbidity in chronic disease. *Journal of chronic diseases* 23:455-468.

Foguet-Boreu, Q., Violan, C., Roso-Llorach, A., Rodriguez-Blanco, T., Pons-Vigues, M., Munoz-Perez, M.A., Pujol-Ribera, E., Valderas, J.M. (2014). Impact of multimorbidity: acute morbidity, area of residency and use of health services across the life span in a region of south Europe. *BMC family practice* 15:55-2296-15-55.

Fortin, M., Bravo, G., Hudon, C., Vanasse, A., Lapointe, L. (2005). Prevalence of multimorbidity among adults seen in family practice. *Annals of family medicine* 3:223-228.

Fortin, M., Stewart, M., Poitras, M.E., Almirall, J., Maddocks, H. (2012). A systematic review of prevalence studies on multimorbidity: toward a more uniform methodology. *Annals of family medicine* 10:142-151.

Fuchs, J., Busch, M., Lange, C., Scheidt-Nave, C. (2012). Prevalence and patterns of morbidity among adults in Germany. Results of the German telephone health interview survey German Health Update (GEDA) 2009. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz* 55:576-586.

Gejman, P.V., Sanders, A.R., Kendler, K.S. (2011). Genetics of schizophrenia: new findings and challenges. *Annual review of genomics and human genetics* 12:121-144.

Genetic Risk and Outcome in Psychosis (GROUP) Investigators. (2011). Evidence that familial liability for psychosis is expressed as differential sensitivity to cannabis: an analysis of patient-sibling and sibling-control pairs. *Archives of General Psychiatry* 68:138-147.

Gottesman, I.I., Gould, T.D. (2003). The endophenotype concept in psychiatry: etymology and strategic intentions. *The American Journal of Psychiatry* 160:636-645.

Ha, N.T., Le, N.H., Khanal, V., Moorin, R. (2015). Multimorbidity and its social determinants among older people in southern provinces, Vietnam. *International journal for equity in health* 14:50-015-0177-8.

Häfner, H. (2003). Gender differences in schizophrenia. *Psychoneuroendocrinology* 28, Supplement 2:17-54.

Hennekens, C.H., Hennekens, A.R., Hollar, D., Casey, D.E. (2005). Schizophrenia and increased risks of cardiovascular disease. *American Heart Journal* 150:1115-1121.

Hoang, U., Goldacre, M.J., Stewart, R. (2013). Avoidable mortality in people with schizophrenia or bipolar disorder in England. *Acta Psychiatrica Scandinavica* 127:195-201.

Hocaoglu, C. (2017). Eating Disorders with Comorbidity Anxiety Disorders. In *Eating Disorders - A Paradigm of the Biopsychosocial Model of Illness*, Jauregui-Lobera, I. (ed), Ch. 06. Rijeka: InTech.

Iacovides, A., Siamouli, M. (2008). Comorbid mental and somatic disorders: an epidemiological perspective. *Current opinion in psychiatry* 21:417-421.

Jacobi, F., Wittchen, H.U., Holting, C., Hofler, M., Pfister, H., Muller, N., Lieb, R. (2004). Prevalence, co-morbidity and correlates of mental disorders in the general population: results from the German Health Interview and Examination Survey (GHS). *Psychological medicine* 34:597-611.

Jarcho, J.M., Mayer, E.A., Jiang, Z.K., Feier, N.A., London, E.D. (2012). Pain, affective symptoms, and cognitive deficits in patients with cerebral dopamine dysfunction. *Pain* 153:744-754.

Keshavan, M.S., Tandon, R., Boutros, N.N., Nasrallah, H.A. (2008). Schizophrenia, "just the facts": what we know in 2008 Part 3: neurobiology. *Schizophrenia research* 106:89-107.

Korver, N., Quee, P.J., Boos, H.B.M., Simons, C.J.P., de Haan, L., Investigators, G. (2012). Genetic Risk and Outcome of Psychosis (GROUP), a multi-site longitudinal cohort study focused on gene-environment interaction: objectives, sample characteristics, recruitment and assessment methods. *International Journal of Methods in Psychiatric Research* 21:205-221.

Lampe, F.C., Walker, M., Lennon, L.T., Whincup, P.H., Ebrahim, S. (1999). Validity of a self-reported history of doctor-diagnosed angina. *Journal of clinical epidemiology* 52:73-81.

Laursen, T.M. (2011). Life expectancy among persons with schizophrenia or bipolar affective disorder. *Schizophrenia research* 131:101-104.

Lawrence, D., Hancock, K.J., Kisely, S. (2013). The gap in life expectancy from preventable physical illness in psychiatric patients in Western Australia: retrospective analysis of population based registers. *BMJ* 346:.

Lawrence, D., Kisely, S. (2010). Inequalities in healthcare provision for people with severe mental illness. *Journal of psychopharmacology (Oxford, England)* 24:61-68.

Manuel, C.M., Rao, P.P., Rebello, P., Safeekh, A.T., Mathai, P.J. (2013). Medical comorbidity in in-patients with psychiatric disorder. *Muller Journal of Medical Science and Research* 4:12-17.

Marengoni, A., Angleman, S., Melis, R., Mangialasche, F., Karp, A., Garmen, A., Meinow, B., Fratiglioni, L. (2011). Aging with multimorbidity: a systematic review of the literature. *Ageing research reviews* 10:430-439.

McGrath, J., Saha, S., Welham, J., El Saadi, O., MacCauley, C., Chant, D. (2004). A systematic review of the incidence of schizophrenia: the distribution of rates and the influence of sex, urbanicity, migrant status and methodology. *BMC Medicine* 2:13.

Molenberghs, G., Verbeke, G. (2006). *Models for Discrete Longitudinal Data*. New York: Springer-Verlag.

Nuyen, J., Schellevis, F.G., Satariano, W.A., Spreeuwenberg, P.M., Birkner, M.D., van den Bos, G.A., Groenewegen, P.P. (2006). Comorbidity was associated with neurologic and psychiatric diseases: a general practice-based controlled study. *Journal of clinical epidemiology* 59:1274-1284.

O'Kelly, S., Smith, S.M., Lane, S., Teljeur, C., O'Dowd, T. (2011). Chronic respiratory disease and multimorbidity: prevalence and impact in a general practice setting. *Respiratory medicine* 105:236-242.

Oreski, I., Jakovljevic, M., Aukst-Margetic, B., Orlic, Z.C., Vuksan-Cusa, B. (2012). Comorbidity and multimorbidity in patients with schizophrenia and bipolar disorder: similarities and differencies. *Psychiatria Danubina* 24:80-85.

Payne, R.A., Abel, G.A., Guthrie, B., Mercer, S.W. (2013). The effect of physical multimorbidity, mental health conditions and socioeconomic deprivation on unplanned admissions to hospital: a retrospective cohort study. *CMAJ : Canadian Medical Association Journal* 185:E221-E228.

Qin, P., Hawton, K., Mortensen, P.B., Webb, R. (2014). Combined effects of physical illness and comorbid psychiatric disorder on risk of suicide in a national population study. *The British Journal of Psychiatry* 204:430-435.

Reilly, S., Olier, I., Planner, C., Doran, T., Reeves, D., Ashcroft, D.M., Gask, L., Kontopantelis, E. (2015). Inequalities in physical comorbidity: a longitudinal comparative cohort study of people with severe mental illness in the UK. *BMJ Open* 5:.

Rizza, A., Kaplan, V., Senn, O., Rosemann, T., Bhend, H., Tandjung, R., FIRE study group. (2012). Age- and gender-related prevalence of multimorbidity in primary care: the Swiss FIRE project. *BMC family practice* 13:113-2296-13-113.

Schafer, I., Hansen, H., Schon, G., Hofels, S., Altiner, A., Dahlhaus, A., Gensichen, J., Riedel-Heller, S., Weyerer, S., Blank, W.A., Konig, H.H., von dem Knesebeck, O., Wegscheider, K., Scherer, M., van den Bussche, H., Wiese, B. (2012). The influence of age, gender and socio-economic status on multimorbidity patterns in primary care. First results from the multicare cohort study. *BMC health services research* 12:89-6963-12-89.

Schneider, F., Kaplan, V., Rodak, R., Battegay, E., Holzer, B. (2012). Prevalence of multimorbidity in medical inpatients. *Swiss medical weekly* 142:w13533.

Seidel, S., Aigner, M., Ossege, M., Pernicka, E., Wildner, B., Sycha, T. (2013). Antipsychotics for acute and chronic pain in adults. *The Cochrane database of systematic reviews* (8):CD004844. doi:CD004844.

Smith, D.J., Langan, J., McLean, G., Guthrie, B., Mercer, S.W. (2013). Schizophrenia is associated with excess multiple physical-health comorbidities but low levels of recorded cardiovascular disease in primary care: cross-sectional study. *BMJ open* 3:10.1136/bmjopen-2013-002808. Print 2013.

Stubbs, B., Koyanagi, A., Veronese, N., Vancampfort, D., Solmi, M., Gaughran, F., Carvalho, A.F., Lally, J., Mitchell, A.J., Mugisha, J., Correll, C.U. (2016). Physical multimorbidity and psychosis: comprehensive cross sectional analysis including 242,952 people across 48 low- and middle-income countries. *BMC medicine* 14:189.

Stubbs, B., Mitchell, A.J., De Hert, M., Correll, C.U., Soundy, A., Stroobants, M., Vancampfort, D. (2014). The prevalence and moderators of clinical pain in people with schizophrenia: a systematic review and large scale meta-analysis. *Schizophrenia research* 160:1-8.

Stubbs, B., Thompson, T., Acaster, S., Vancampfort, D., Gaughran, F., Correll, C.U. (2015). Decreased pain sensitivity among people with schizophrenia: a meta-analysis of experimental pain induction studies. *Pain* 156:2121-2131.

Taylor, A.W., Price, K., Gill, T.K., Adams, R., Pilkington, R., Carrangis, N., Shi, Z., Wilson, D. (2010). Multimorbidity - not just an older person's issue. Results from an Australian biomedical study. *BMC public health* 10:718-2458-10-718.

Tomasdottir, M.O., Getz, L., Sigurdsson, J.A., Petursson, H., Kirkengen, A.L., Krokstad, S., McEwen, B., Hetlevik, I. (2013). Co- and multi-morbidity patterns in an unselected Norwegian population: cross-sectional analysis

based on the HUNT Study and theoretical reflections concerning basic medical models. *European Journal for Person Centered Healthcare* 2:335-345.

van den Akker, M., Buntinx, F., Metsemakers, J.F., Roos, S., Knottnerus, J.A. (1998). Multimorbidity in general practice: prevalence, incidence, and determinants of co-occurring chronic and recurrent diseases. *Journal of clinical epidemiology* 51:367-375.

van Oostrom, S.H., Picavet, H.S., van Gelder, B.M., Lemmens, L.C., Hoeymans, N., van Dijk, C.E., Verheij, R.A., Schellevis, F.G., Baan, C.A. (2012). Multimorbidity and comorbidity in the Dutch population - data from general practices. *BMC public health* 12:715-2458-12-715.

van Winkel, R., De Hert, M., Van Eyck, D., Hanssens, L., Wampers, M., Scheen, A., Peuskens, J. (2006). Screening for diabetes and other metabolic abnormalities in patients with schizophrenia and schizoaffective disorder: evaluation of incidence and screening methods. *The Journal of clinical psychiatry* 67:1493-1500.

Vancampfort, D., Correll, C.U., Galling, B., Probst, M., De Hert, M., Ward, P.B., Rosenbaum, S., Gaughran, F., Lally, J., Stubbs, B. (2016). Diabetes mellitus in people with schizophrenia, bipolar disorder and major depressive disorder: a systematic review and large scale meta-analysis. *World psychiatry: official journal of the World Psychiatric Association (WPA)* 15:166-174.

Vancampfort, D., Stubbs, B., Mitchell, A.J., De Hert, M., Wampers, M., Ward, P.B., Rosenbaum, S., Correll, C.U. (2015). Risk of metabolic syndrome and its components in people with schizophrenia and related psychotic disorders, bipolar disorder and major depressive disorder: a systematic review and meta-analysis. *World psychiatry: official journal of the World Psychiatric Association (WPA)* 14:339-347.

Verhage, F. (1964). *Intelligentie en leeftijd: Onderzoek bij Nederlanders van twaalf tot zevenenzeventig jaar*. Assen: Van Gorcum.

Vos, H.M., Bor, H.H., Rangelrooij-Minkels, M.J., Schellevis, F.G., Lagro-Janssen, A.L. (2013). Multimorbidity in older women: the negative impact of specific combinations of chronic conditions on self-rated health. *The European journal of general practice* 19:117-122.

Walker, E.R., McGee, R.E., Druss, B.G. (2015). Mortality in mental disorders and global disease burden implications: a systematic review and meta-analysis. *JAMA psychiatry* 72:334-341.

Wechsler, D., Coalson, D.L., Raiford, S.E. (2008). *Wechsler Adult Intelligence Scale: Technical and interpretive manual*. San Antonio: TX: Pearson.

Willadsen, T.G., Bebe, A., Koster-Rasmussen, R., Jarbol, D.E., Guassora, A.D., Waldorff, F.B., Reventlow, S., Olivarius Nde, F. (2016). The role of diseases, risk factors and symptoms in the definition of multimorbidity - a systematic review. *Scandinavian journal of primary health care* 34:112-121.

Supplementary Materials

Supplementary Methods

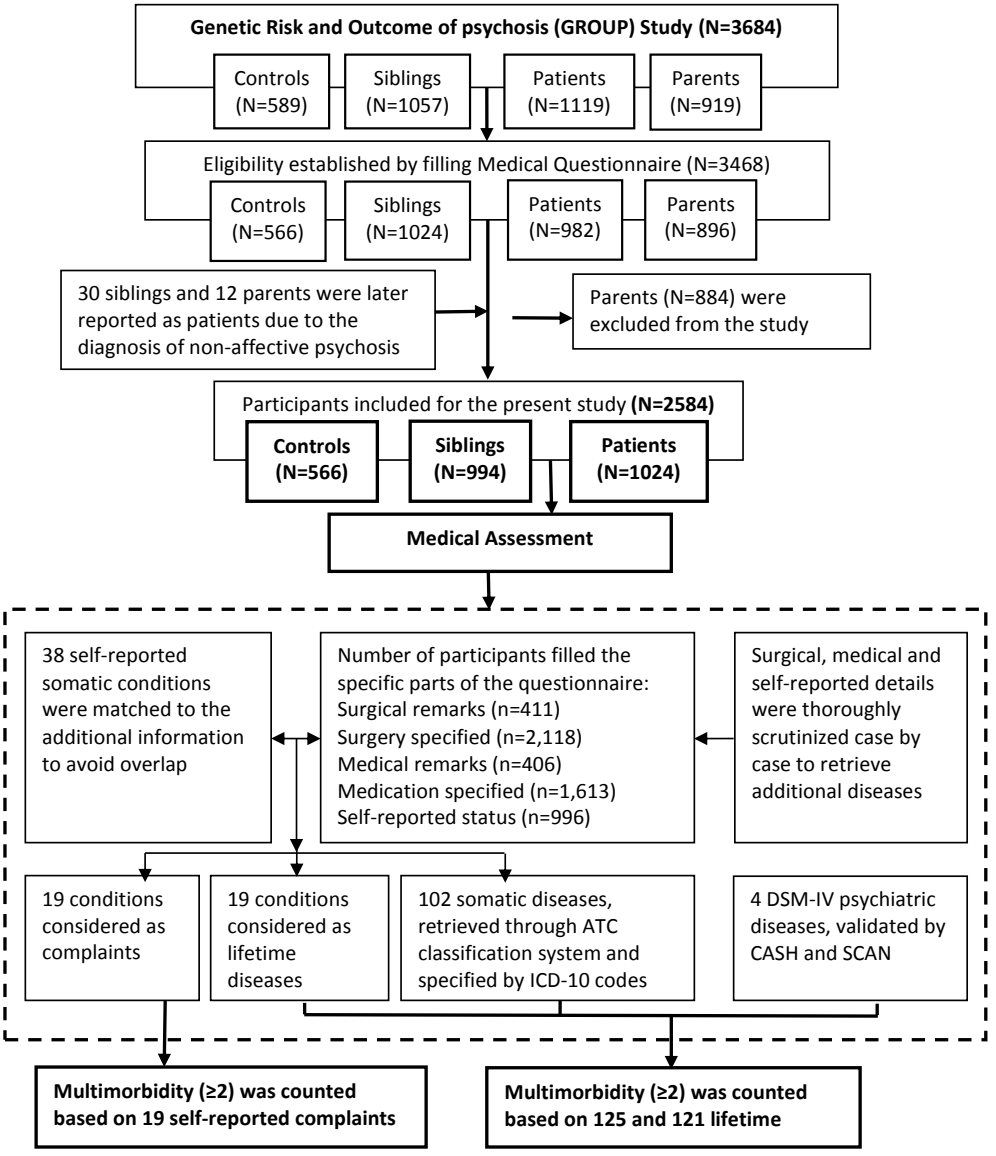
In Genetic Risk and Outcome of Psychosis (GROUP) project, patients were identified through clinicians working in regional psychosis departments or academic centers, whose caseload was screened for inclusion criteria. Subsequently, a group of patients presenting consecutively at these services either as outpatients or inpatients were recruited for the study from 2007 to 2008. Persons identified as potentially eligible were given a detailed explanation of the study procedures and were asked for informed consent for detailed assessment and for contacting their first-degree family members (brothers, sisters, parents). Controls were selected through a system of random mailings to addresses in the catchment areas of the cases (Korver et al., 2012).

The selection of the sample is depicted in Supplementary Figure S1. We followed sample selection procedure for 3684 participants, case by case, to retrieve information regarding lifetime diseases and complaints. Among them, 3468 participants filled in the Medical Questionnaire (MQ). We thoroughly scrutinized the MQ file, containing medical disease related data collected through a 'Medische Vragenlijst (medical questionnaire)'. The checklist section contained information on 38 medical conditions in the form of 'present/absent'. The narrative section provided a detailed medical, surgical, and medication history, as well as remarks on checklist status. We further retrieved compatible disease relevant information from narrative questions. Additionally, medication history was reviewed by the Anatomical Therapeutic Chemical (ATC) classification system and furthers the ICD-10 to obtain additional information to confirm evidence of the specific entities of the diseases derived from the medical questionnaire and registered diagnoses of the participants. According to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) codes, and information either from the Schedule for Clinical Assessment in Neuropsychiatry (SCAN) or the Comprehensive Assessment of Symptoms and History (CASH), five groups of psychiatric disorders (psychotic: including schizophrenia, schizoaffective and delusional; mood; anxiety and/or eating; substance, and others) were formed regarding patients, siblings and controls, respectively, at baseline. After gathering all information, we performed crosschecking across data domains to avoid duplication.

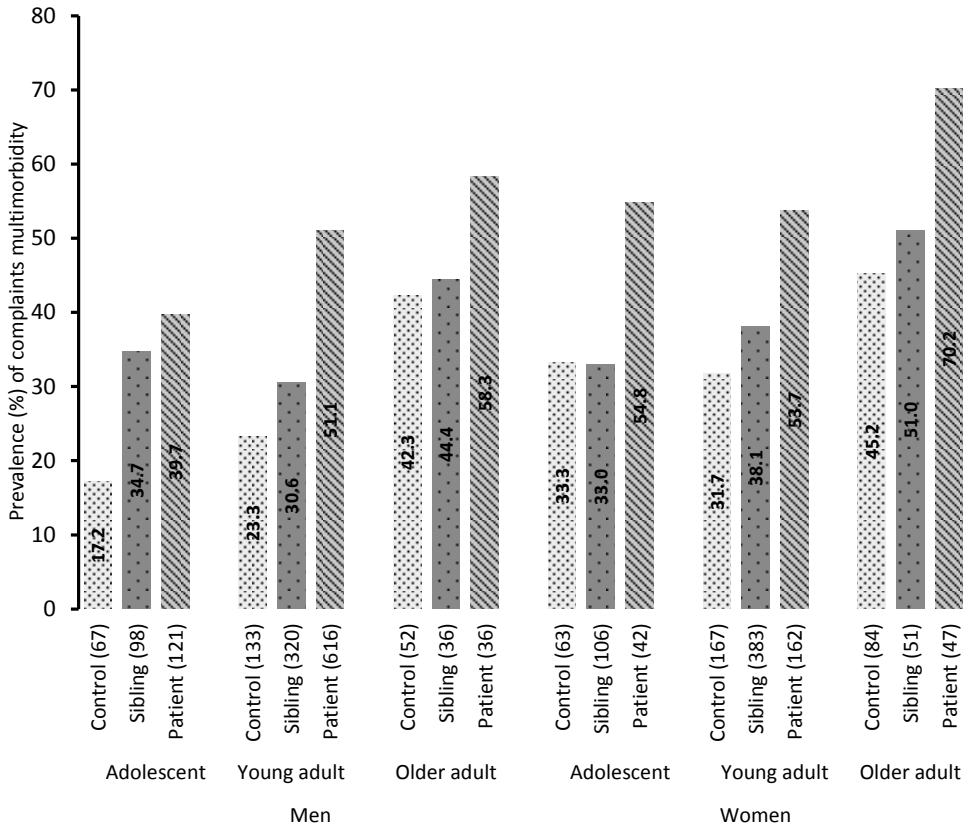
Supplementary Results

The observed prevalence of multimorbidity of complaints, physical health and lifetime diseases across familial liability groups by gender and age groups simultaneously are demonstrated in Supplementary Figures S2-S4. The comparisons between patient-control, sibling-control and patient-sibling on individual complaints are presented in Supplementary Table S1. More specifically, presence of the complaints angina pectoris, palpitation, short of breath, and dizziness were significantly different across the sibling-control, patient-control and patient-sibling groups (Supplementary Table S1). An overview of the group frequencies of all individual diseases is presented in Supplementary Table S2. However, a meaningful statistical analysis was not possible for the individual lifetime diseases due to very low numbers of subjects per group. Instead, the domains of lifetime diseases were compared across different familial liability subgroups, and are presented in Supplementary

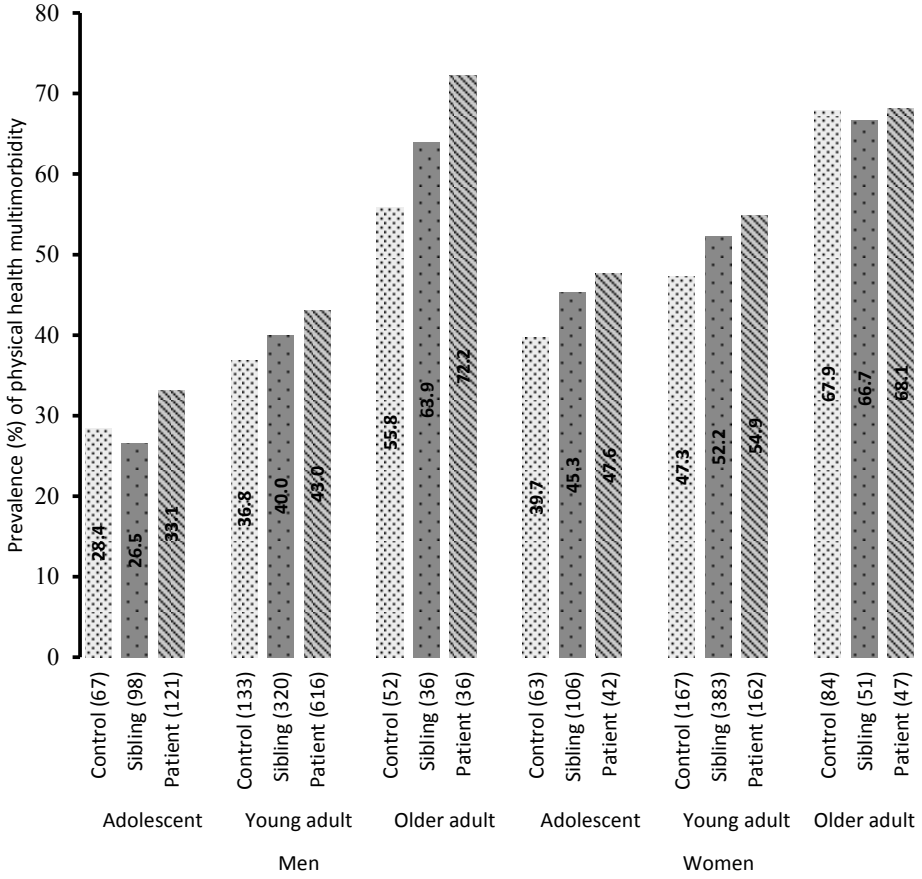
Table S3. Several domains of disease, such as diseases of the nervous system (35%), congenital mal/del-formations, and chromosomal abnormalities (12%), observed a *priori* in schizophrenia patients, also showed a significantly higher frequency compared to siblings and controls (Supplementary Table S3). The most prevalent lifetime diseases in patients, siblings and controls were concussion (24.4%, 15.3% and 19.3%), eczema (18.9%, 16.3% and 18.0%), migraine (8.8%, 10.1% and 10.2%), tonsillitis (9.3%, 12.5% and 8.1%), congenital defects (12.0%, 8.8% and 6.4%) and mood disorders (5.5%, 11.9% and 9.2%). The presence of concussion, mood disorders and congenital defects differed significantly between patients, and controls and siblings. Other diseases had an occurrence of less than 5% across the familial liability groups (Supplementary Table S4). Other schizophrenia studies showed different commonly co-occurring diseases such as hypertension, impaired glucose tolerance, diabetes, cardiovascular diseases, chronic infections and lung diseases, resulting overall in premature death, and particularly at a younger age (Bresee et al., 2010; Bushe and Holt, 2004; De Hert et al., 2009; Hennekens et al., 2005; Iacovides and Siamouli, 2008; van Winkel et al., 2006). Their results suggested a different profile of multimorbid diseases in younger patients with schizophrenia, who were already suffering from multiple medical conditions upon onset of their disease.



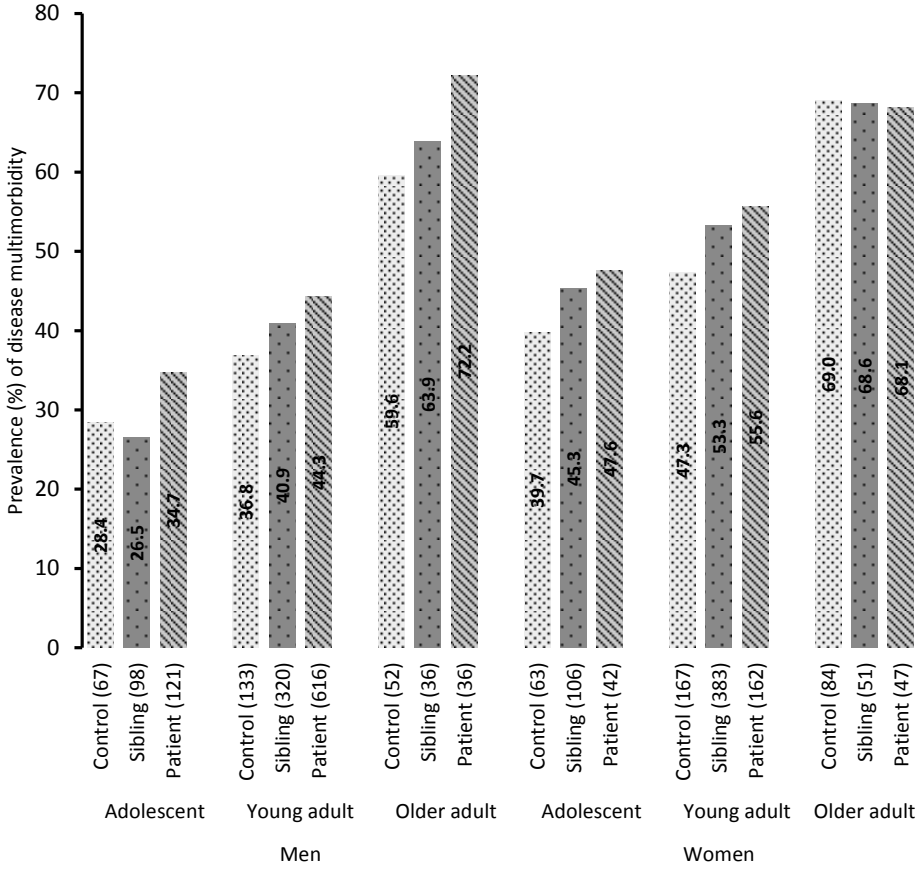
Supplementary Figure S1: Flowchart of inclusion of participants (GROUP version 3.2) and retrieval of information on complaints and diseases.



Supplementary Figure S2: Prevalence of multimorbidity of complaints across familial liability groups by gender and age groups simultaneously.



Supplementary Figure S3: Prevalence of multimorbidity of physical health diseases across familial liability groups by gender and age groups simultaneously.



Supplementary Figure S4: Prevalence of multimorbidity of lifetime diseases across familial liability groups by gender and age groups simultaneously.

Supplementary Table S1: Relative frequencies (percentages) and patient-control, sibling-control and patient-sibling comparisons of complaints.

Name of Complaint	Familial liability group					
	Control	Sibling vs. Control		Patient vs. Control		Patient vs. Sibling
	N (%)	N (%)	P-value	N (%)	P-value	P-value
1. Angina pectoris	16 (2.8)	48 (4.8)	0.044	130 (12.7)	<0.001	<0.001
2. Palpitation	27 (4.8)	84 (8.5)	0.008	188 (18.4)	<0.001	<0.001
3. Oedema	16 (2.8)	38 (3.8)	0.307	50 (4.9)	0.056	0.253
4. Short of breath	15 (2.7)	73 (7.3)	<0.001	187 (18.3)	<0.001	<0.001
5. Hyperventilation	28 (4.9)	64 (6.4)	0.230	124 (12.1)	<0.001	<0.001
6. Dizziness	41 (7.2)	115 (11.6)	0.007	238 (23.2)	<0.001	<0.001
7. Perspiration	32 (5.7)	72 (7.2)	0.242	172 (16.8)	<0.001	<0.001
8. Pyrosis	40 (7.1)	81 (8.1)	0.449	182 (17.8)	<0.001	<0.001
9. Food allergy	51 (9.0)	84 (8.5)	0.708	119 (11.6)	0.106	0.017
10. Problematic bowel movement	57 (10.1)	111 (11.2)	0.514	163 (15.9)	0.001	0.002
11. Jaundice	7 (1.2)	10 (1.0)	0.674	10 (1.0)	0.630	0.947
12. Paralysis	7 (1.2)	8 (0.8)	0.405	22 (2.1)	0.223	0.016
13. Pain in joints or muscles	99 (17.5)	182 (18.3)	0.676	220 (21.5)	0.062	0.081
14. Limited movement of joints (stiffness)	30 (5.3)	53 (5.3)	0.982	69 (6.7)	0.286	0.198
15. Loss of hearing	32 (5.7)	51 (5.1)	0.606	89 (8.7)	0.035	0.001
16. Loss of smell	15 (2.7)	32 (3.2)	0.555	82 (8.0)	<0.001	<0.001
17. ENT disorder (except loss of hearing)	39 (6.9)	82 (8.2)	0.338	92 (9.0)	0.147	0.551
18. Allergy	118 (20.8)	166 (16.7)	0.047	163 (15.9)	0.015	0.625
19. Dermatology disorder (except allergy)	41 (7.2)	95 (9.6)	0.143	87 (8.5)	0.396	0.436

Supplementary Table S2: Relative frequencies (percentages) of all lifetime diseases for controls, siblings and patients.

Name of lifetime diseases	Familial liability group		
	Control (N=566)	Sibling (N=994)	Patient (N=1024)
Mental and behavioral disorders			
Psychotic disorders (Schizophrenia, schizoaffective and delusional)	…	…	1024 (39.6)*
1. Mood Disorder	52 (9.2)	118 (11.9)	56 (5.5)
2. Anxiety and/or Eating Disorder	1 (0.2)	10 (1.0)	6 (0.6)
3. Substance Disorder	0	4 (0.4)	21 (2.1)
4. Other Disorders[a]	10 (1.8)	19 (1.9)	20 (2.0)
Diseases of the circulatory system			
5. Ischemic Heart Disease	2 (0.4)	4 (0.4)	7 (0.7)
6. Aneurysm	0	1 (0.1)	0
7. Myocarditis	0	0	0
8. Cardiomyopathy	0	0	1 (0.1)
9. Kawasaki Disease	0	1 (0.1)	0
10. Heart Failure	0	0	0
11. Arrhythmia	3 (0.5)	1 (0.1)	1 (0.1)
12. Heart Valve Disease	2 (0.4)	1 (0.1)	2 (0.2)
13. Hypertension	5 (0.9)	8 (0.8)	12 (1.2)
14. Hypotension	1 (0.2)	4 (0.4)	0
15. Stroke	1 (0.2)	1 (0.1)	6 (0.6)
16. Transient Ischemic Attack	1 (0.2)	2 (0.2)	2 (0.2)
17. Deep Vein Thrombosis	0	1 (0.1)	1 (0.1)
18. Pulmonary Embolism	0	2 (0.2)	1 (0.1)
19. Phlebitis	0	1 (0.2)	0
20. Hemorrhoids/Varices	0	2 (0.2)	0
21. Varicosity	7 (1.2)	7 (0.7)	2 (0.2)
Endocrine, nutritional and metabolic diseases			
22. Dyslipidaemia	5 (0.9)	5 (0.5)	6 (0.6)
23. Obesity	23 (4.1)	38 (3.8)	62 (6.1)
24. Diabetes	3 (0.5)	8 (0.8)	16 (1.6)
25. Hypothyroidism	3 (0.5)	9 (0.9)	8 (0.8)
26. Hyperthyroidism	1 (0.2)	0	3 (0.3)
27. Thyroiditis (not specified)	4 (0.7)	7 (0.7)	10 (1.0)
28. Goiter	2 (0.4)	0	0
29. Addison's Disease	0	0	1 (0.1)
Diseases of the respiratory system			
30. Chronic Obstructive Pulmonary Disease	26 (4.6)	52 (5.2)	74 (7.2)
31. Asthma	10 (1.8)	17 (1.7)	5 (0.5)
32. Pneumothorax	0	2 (0.2)	1 (0.1)
33. Bronchitis	0	1 (0.1)	1 (0.1)
34. Pneumonia	0	0	1 (0.1)
35. Sleep Apnea	0	0	1 (0.1)
36. Sarcoidosis	1 (0.2)	0	0
37. Nasal Polyps	3 (0.9)	8 (0.8)	4 (0.4)
38. Deviated Nasal Septum	5 (0.9)	7 (0.7)	6 (0.6)
39. Epistaxis	0	1 (0.1)	2 (0.2)
40. Chronic Sinusitis	0	4 (0.4)	2 (0.2)
41. Tonsillitis	46 (8.1)	124 (12.5)	95 (9.3)
42. Chronic Laryngitis	1 (0.2)	1 (0.1)	0
43. Hay Fever	49 (8.7)	65 (6.5)	51 (5.0)

Neoplasms			
44. Cancer (not specified)	4 (0.7)	9 (0.9)	9 (0.9)
Diseases of the musculoskeletal system and connective tissue			
45. Rheumatoid Arthritis	2 (0.4)	2 (0.2)	0
46. Systemic lupus Erythematosus	1 (0.2)	0	0
47. Multiple Sclerosis	0	0	0
48. Psoriatic Arthritis	0	0	0
49. Osteoarthritis	3 (0.5)	7(0.7)	12 (1.2)
50. Frozen Shoulder	0	0	0
51. Gout	1 (0.2)	0	0
52. Fibromyalgia	4 (0.7)	6 (0.6)	3 (0.3)
53. Back & Neck Pain	16 (2.8)	16 (1.6)	9 (0.9)
54. Scoliosis & Kyphosis	4 (0.7)	6 (0.6)	1 (0.1)
55. Carpal Tunnel Syndrome	2 (0.4)	2 (0.2)	2 (0.2)
56. Osteoporosis	2 (0.4)	1 (0.1)	0
57. Osteomyelitis	1 (0.2)	0	0
Diseases of the digestive system			
58. Chronic Gastritis	6 (1.1)	10 (1.0)	16 (1.6)
59. Helicobacter Pylori	0	1 (0.1)	0
60. Lactose Intolerance	4 (0.7)	7 (0.7)	4 (0.4)
61. Diverticulitis	0	0	0
62. Celiac Disease	0	1 (0.1)	0
63. Diarrhea	0	4 (0.4)	3 (0.3)
64. Constipation	2 (0.4)	5 (0.5)	10 (1.0)
65. Peritonitis	0	0	1 (0.1)
66. Pancreatitis	0	0	0
67. Irritable Bowel Syndrome	4 (0.7)	8 (0.8)	2 (0.2)
68. Inflammatory Bowel Disorder	7 (1.2)	4 (0.4)	1 (0.1)
69. Gall Bladder Diseases	6 (1.1)	9 (0.9)	13 (1.3)
70. Diseases of Appendix	27 (4.8)	60 (6.0)	47 (4.6)
71. Hernia	22 (3.9)	32 (3.2)	34 (3.3)
72. Anal Fistula	2 (0.4)	3 (0.3)	3 (0.3)
73. Teeth Abnormalities	12 (2.1)	16 (1.6)	12 (1.2)
Pregnancy, childbirth and the puerperium			
74. Abortion/Recurrent Abortion	1 (0.2)	4 (0.4)	6 (0.6)
75. Ectopic Pregnancy	2 (0.4)	2 (0.2)	0
76. Other Obstetric Complications	1 (0.2)	2 (0.2)	0
Diseases of the genitourinary system			
77. Diseases of Genital Organ	3 (0.5)	5 (0.5)	3 (0.3)
78. Infertility	2 (0.4)	2 (0.2)	1 (0.1)
79. Endometriosis	0	3 (0.3)	1 (0.1)
80. Menopausal Syndrome	1 (0.2)	0	0
81. Prolapse	0	1 (0.1)	0
82. Menstrual Abnormalities	1 (0.2)	0	1 (0.1)
83. Pelvic Inflammatory Disease	0	1 (0.1)	1 (0.1)
84. Chronic Cystitis (>3/years)	18 (3.2)	40 (4.0)	31 (3.0)
85. Kidney Stone	11 (1.9)	13 (1.3)	8 (0.8)
86. Chronic Kidney Disease	0	1 (0.1)	0
87. Pyelonephritis	1 (0.2)	1 (0.1)	0
88. Hydrocele	1 (0.2)	1 (0.1)	0
89. Benign Prostatic Hyperplasia	1 (0.2)	1 (0.1)	0
90. Epididymitis	1 (0.2)	0	0
Certain infectious and parasitic diseases			
91. Tuberculosis lifetime	5 (0.9)	3 (0.3)	10 (1.0)
92. Viral Hepatitis	2 (0.4)	2 (0.2)	3 (0.3)
93. Hepatitis (not specified)	1 (0.2)	0 (0.0)	4 (0.4)

94. Cytomegalovirus	1 (0.2)	1 (0.1)	0
95. Infectious Mononucleosis	0	7 (0.7)	0
96. Tropical Diseases (not specified)	1 (0.2)	8 (0.8)	2 (0.2)
97. Other Bacterial Infection	1 (0.2)	1 (0.1)	3 (0.3)
98. Infectious Diseases (not specified)	15 (2.7)	22 (2.2)	38 (3.7)
99. Parasitic Infestation	0	3 (0.3)	2 (0.2)
100. Childhood Infectious Diseases	1(0.2)	1 (0.1)	1(0.1)
Diseases of the skin and subcutaneous tissue			
101. Eczema	102 (8.0)	162 (6.3)	194 (8.9)
102. Acne	10 (1.8)	6 (0.6)	8 (0.8)
103. Lichen Planus	0	0	1 (0.1)
104. Psoriasis	3 (0.2)	5 (0.5)	4 (0.4)
105. Pigment Disorders	0	0	2 (0.2)
Diseases of the eye and adnexa			
106. Vision Reduction	18 (3.2)	27 (2.7)	18 (1.8)
107. Lazy Eye	3 (0.5)	5 (0.5)	3 (0.3)
108. Coordination Abnormalities	2 (0.4)	1 (0.1)	4 (0.4)
109. Glaucoma	1 (0.2)	1 (0.1)	0
110. Others Eye Disorders	0	3 (0.3)	1 (0.1)
Diseases of the ear and mastoid process			
111. Hearing Impairment (diagnosed)	4 (0.7)	10 (1.0)	8 (0.8)
112. Chronic Suppurative Otitis Media	16 (2.8)	29 (2.9)	19 (1.9)
113. Cholesteatoma	1 (0.2)	1 (0.1)	0
114. Otosclerosis	1 (0.2)	0	0
Diseases of the nervous system (+ due to injury)			
115. Epilepsy	7 (1.2)	9 (0.9)	24 (2.3)
116. Concussion	109 (19.3)	152 (15.3)	250 (24.4)
117. Meningitis	5 (0.9)	7 (0.7)	13 (1.3)
118. Migraine	58 (10.2)	100 (10.1)	90 (8.8)
119. Insomnia	0	1 (0.1)	1 (0.1)
120. Chronic Fatigue Syndrome	1 (0.2)	0	0
121. Impaired Smell	0	0	6 (0.6)
Diseases of blood & blood-forming organs and certain disorders involving the immune mechanism			
122. Anemia	2 (0.4)	7 (0.7)	1 (0.1)
Congenital malformations, deformations and chromosomal abnormalities			
123. Congenital Defect	36 (6.4)	87 (8.8)	123 (12.0)
124. Sex-linked Diseases	0	0	0
125. Sexual Dysfunction	0	2 (0.2)	0

*Out of 2584 subjects, [a]Other disorders= Adjustment disorder, bereavement, personality disorder, Autism, Deferred.

Supplementary Table S3: Relative frequencies (percentages) and patient-control, sibling-control and patient-sibling comparisons per domain for the lifetime diseases.

Domain	Familial liability group					
	Control	Sibling vs. Control		Patient vs. Control		Patient vs. Sibling
	N (%)	N (%)	P-value	N (%)	P-value	P-value
1. Mental and behavioral disorders	61 (10.8)	144 (14.5)	0.038	103 (10.1)	0.659	0.003
2. Diseases of circulatory system	22 (3.9)	35 (3.5)	0.720	35 (3.4)	0.627	0.882
3. Endocrine, nutritional and metabolic diseases	37 (6.5)	64 (6.4)	0.937	95 (9.3)	0.066	0.015
4. Diseases of respiratory system	179 (31.6)	348 (35.0)	0.162	298 (29.1)	0.292	0.003
5. Neoplasms	4 (0.7)	9 (0.9)	0.679	9 (0.9)	0.716	0.950
6. Diseases of musculoskeletal system and connective tissue	34 (6.0)	39 (3.9)	0.067	27 (2.6)	0.001	0.950
7. Diseases of digestive system	80 (14.1)	138 (13.9)	0.865	125 (12.2)	0.290	0.089
8. Pregnancy, childbirth and the Puerperium	4 (0.7)	8 (0.8)	0.831	6 (0.6)	0.771	0.265
9. Diseases of genitourinary system	39 (6.9)	64 (6.4)	0.730	44 (4.3)	0.028	0.555
10. Certain infectious and parasitic diseases	24 (4.2)	44 (4.4)	0.833	58 (5.7)	0.226	0.034
11. Diseases of the skin and subcutaneous tissue	114 (20.1)	175 (17.6)	0.191	208 (20.3)	0.979	0.223
12. Diseases of the eye and adnexa	24 (4.2)	37 (3.7)	0.616	25 (2.4)	0.051	0.105
13. Diseases of the ear and mastoid process	22 (3.9)	40 (4.0)	0.920	26 (2.5)	0.145	0.089
14. Diseases of the nervous system (+ due to injury)	169 (29.9)	264 (26.6)	0.165	358 (35.0)	0.041	0.067
15. Diseases of blood & blood-forming organs and certain disorders involving the immune mechanism	2 (0.4)	7 (0.7)	0.389	1 (0.1)	0.293	<0.001
16. Congenital mal/de-formations and chromosomal abnormalities	36 (6.4)	87 (8.8)	0.096	123 (12.0)	<0.001	0.064

Note: The 16 domains were based on 125 lifetime diseases.

Supplementary Table S4: Top 10 lifetime diseases ranked by controls.

Comorbid Diseases	Familial liability group					
	Control	Sibling vs. Control		Patient vs. Control		Patient vs. Sibling
	%	%	P-value	%	P-value	P-value
1. Concussion	19.3	15.3	0.046	24.4	0.019	<0.001
2. Eczema	18.0	16.3	0.362	18.9	0.641	0.097
3. Migraine	10.2	10.1	0.916	8.8	0.314	0.280
4. Mood Disorders	9.2	11.9	0.105	5.5	0.006	<0.001
5. Hay Fever	8.7	6.5	0.745	5.0	0.069	0.009
6. Tonsillitis	8.1	12.5	0.013	9.3	0.424	0.029
7. Congenital defect	6.4	8.8	0.096	12.0	0.001	0.016
8. Appendicitis	4.8	6.0	0.297	4.6	0.862	0.135
9. COPD[a]	4.6	5.2	0.572	7.2	0.045	0.066
10. Obesity	4.1	3.8	0.855	6.1	0.088	0.014

Note: Table presents the percentages; [a]COPD=Chronic Obstructive Pulmonary Disease

CHAPTER 7

Factors contributing to the duration of untreated psychosis

Sanne Y. Apeldoorn

Bouke Sterk

Edwin R. van den Heuvel

Robert A. Schoevers

Md. Atiqul Islam

GROUP Investigators

Schizophrenia Research 2014, 158: 76-81

Abstract

Background: Shortening the duration of untreated psychosis (DUP) - with the aim of improving the prognosis of psychotic disorders - requires an understanding of the causes of treatment delay. Current findings concerning several candidate risk factors of a longer DUP are inconsistent. Our aim was to identify factors contributing to DUP in a large sample that represents the treated prevalence of non-affective psychotic disorders.

Method: Patients with a non-affective psychotic disorder were recruited from mental health care institutes from 2004 to 2008. Of the 1120 patients enrolled, 852 could be included in the present analysis. Examined candidate factors were gender, educational level, migration status, premorbid adjustment and age at onset of the psychotic disorder. DUP was divided into five ordinal categories: less than one month, one month to three months, three months to six months, six months to twelve months and twelve months and over. An ordinal logistic regression analysis was used to identify the risk factors of a longer DUP.

Results: Median DUP was less than one month (IQR 2). The factors migration status ($p = 0.028$), age at onset of the psychotic disorder ($p = 0.003$) and gender ($p = 0.034$) were significantly associated with DUP in our analysis.

Conclusion: First generation immigrant patients, patients with an early onset of their psychotic disorder and male patients seem at risk of a longer DUP. These findings can assist in designing specific interventions to shorten treatment delay.

Keywords: duration of untreated psychosis; DUP; migration status; age at onset; gender

1. Introduction

The duration of untreated psychosis (DUP) is defined as the time from the emergence of the first psychotic episode to the initiation of adequate treatment. DUP can last days, months or even years (Marshall et al., 2005). A longer DUP is associated with worse short-term (Marshall et al., 2005; Perkins et al., 2005) and long-term outcomes (Bottlender et al., 2003; Crumlish et al., 2009; Boonstra et al., 2012a). The potential of DUP being modifiable raises the possibility of improving outcome by shortening DUP. In designing interventions to shorten DUP, it is important to identify factors contributing to DUP.

Factors previously associated with a longer DUP include stigma-related concerns (Corrigan, 2004; Tanskanen et al., 2011), an insidious mode of onset (Morgan et al., 2006; Compton et al., 2008) and a diagno-sis of non-affective psychosis compared with affective psychosis (Morgan et al., 2006; Bechard-Evans et al., 2007; Schimmelmann et al., 2008).

Inconsistent results have been reported for the association between DUP and several other factors. Concerning gender, even though studies continue to show that men have (a tendency for) a longer DUP compared with women (Chang et al., 2011; Fridgen et al., 2012), a review could not confirm the association (Cascio et al., 2012). Also inconsistent are the findings with respect to the association between DUP and educational level: longer DUP was found to be associated with a higher level (Chong et al., 2005), a lower level (Verdoux et al., 1998) and not with educational level at all (Morgan et al., 2006; Bechard-Evans et al., 2007; Compton et al., 2008, 2011). Most studies did not find an association between DUP and ethnicity (Anderson et al., 2013). Interestingly however, three recent studies reported an association between DUP and migration status (Sterk et al., 2010; Boonstra et al., 2012b; Nerhus et al., 2013). Furthermore, inconsistent results have been reported with respect to the association between DUP and overall premorbid adjustment (Chen et al., 2005; versus Harrigan et al., 2003; Schimmelmann et al., 2008) and the association between DUP and age at onset of the psychotic disorder (Bechard-Evans et al., 2007; Schimmelmann et al., 2008; versus Drake et al., 2000; Morgan et al., 2006). Notably, many previous studies examining DUP had relatively small sample sizes, a mixed sample of patients with affective and non-affective psychotic disorders and a substantial variation in definition of DUP.

Given the importance of knowledge concerning the factors associated with DUP, the inconsistency of previous findings and the limitations of previous research, the association between DUP and candidate risk factors needs further elucidation. The aim of this study was to identify risk factors of a longer DUP in a large sample that represents the treated prevalence of non-affective psychotic disorders. Specifically, we aimed to test the hypothesis that being an immigrant, having a poor premorbid adjustment and having an earlier age at onset of the psychotic disorder is associated with a longer DUP. Furthermore, we hypothesized that gender and educational level were not associated with DUP.

2. Methods

2.1. Study design and population

Data were extracted from the baseline assessments of a longitudinal, multi-site, naturalistic cohort study: the Genetic Risk and Outcome of Psychosis (GROUP) study (data release 3.02). Patients were

recruited from mental health care institutes in selected representative geographical areas in the Netherlands and Belgium. They were identified through clinicians, whose caseloads were screened for inclusion criteria. Subsequently, patients presenting consecutively at these services either as outpatients or inpatients were recruited. Inclusion criteria for patients were: 1) age range of 16 to 50 years; 2) diagnosis of a non-affective psychotic disorder, according to the Diagnostic and Statistical Manual of Mental Disorders Fourth Edition (DSM-IV) criteria; and 3) good command of the Dutch language. The GROUP study protocol was approved centrally by the Ethical Review Board of the University Medical Centre Utrecht, and subsequently by local review boards of each participating institute. Informed consent was obtained from all participants after complete description of the study and before the start of the first assessment. Detailed information about the GROUP study is published elsewhere (Korver et al., 2012).

The GROUP sample consisted of 1120 patients, of which 852 could be included in the present study. Reasons for and numbers of exclusion were the following: 1) 19 patients were excluded because of a final diagnosis other than a non-affective psychotic disorder; 2) 226 patients were excluded because DUP could not be calculated, as data were incomplete; and 3) 23 patients were excluded because their calculated DUP was longer than the recorded duration of the first psychotic episode, meaning data were incorrect.

2.2. Definitions and measures

To establish the DSM-IV diagnosis of a non-affective psychotic disorder two structured diagnostic instruments were used, in accordance with the standard practice in the study sites: the Comprehensive Assessment of Symptoms and History (CASH) (Andreasen et al., 1992) and the Schedules for Clinical Assessment for Neuropsychiatry (SCAN 2.1) (Wing et al., 1990). All raters had completed training in the instruments and diagnostic consensus was achieved in the presence of an independent psychiatrist.

Information regarding DUP was assessed with the Life Chart Schedule (LCS) (Sartorius et al., 1996) by clinical trained interviewers. DUP was defined as the number of months from the onset of the first psychotic episode to the initiation of appropriate treatment for this episode. The first psychotic episode was defined as the first period in which hallucinations, delusions and/or clear disorganized speech or thinking were present for at least one week, according to criteria by the CASH or SCAN. Appropriate treatment, the end-point for DUP, was defined as the use of antipsychotic medication and/or regular treatment contact with a mental health professional for psychosis. The starting month and year of receiving medication or initiation of treatment contact was noted – whichever started first. When treatment was started before the onset of the first psychotic episode, this resulted in a negative DUP. This can happen in case treatment is started during the prodromal phase. These negative values were truncated to zero values.

The following variables were considered as candidate factors contributing to DUP: gender, educational level, migration status, premorbid adjustment and age at onset of the psychotic disorder. Educational level was based on a subdivision by Verhage (Verhage, 1964) and ranged from zero (no education) to eight (university degree). Together the subdivisions of lower, higher and pre-university

secondary education bear resemblance with internationally well-known secondary education. The three types of vocational education should be regarded as "universities of professional education".

Migration status was defined as follows: when a patient and at least one of the parents were born abroad, the patient was classified as a first generation immigrant. When a patient was born in the Netherlands or Belgium and at least one of the parents was born abroad, the patient was classified as a second generation immigrant. All other patients were considered as natives.

The Premorbid Adjustment Scale (PAS) (Cannon-Spoor et al., 1982) was used to determine premorbid adjustment. PAS is designed to evaluate the levels of functioning at several periods of a subject's life, before the onset of the psychotic disorder. It covers sociability and withdrawal, peer relationships, scholastic performance, adaption to school and capacity to establish socio-sexual relationships. For analyses a PAS overall score was used, calculated by averaging the period scores - before onset of the psychotic disorder - per patient. Age at onset of the psychotic disorder was, like information regarding DUP, assessed by using the LCS.

2.3. Statistical analysis

Patient's characteristics were summarized by using descriptive statistics. Differences between the included and excluded patients were tested using Mann-Whitney tests, chi-squared tests and Fisher's exact tests as appropriate. Due to the very high positive skewed distribution of DUP, it was necessary to convert DUP into categories prior to statistical analysis. Because there are no agreed-on cutoff points (Marshall et al., 2005), categorization was based on a combination of cutoff points selected in two previous studies (Harrigan et al., 2003; Chang et al., 2012). DUP was converted into a set of five ordinal categories: less than one month; one month to three months; three months to six months; six months to twelve months; and twelve months and over. An ordinal logistic regression analysis was used to analyze the data.

The full model included all five preselected candidate risk factors, based on literature. We then applied a backward selection procedure to come up to our final model, by eliminating candidate risk factors one by one when p-values for all levels from Type 3 tests were larger than or equal to 0.05. The proportional odds assumption was checked for the final model.

Analyses were conducted using Statistical Analysis System (SAS Institute Inc., Cary, NC). The level of significance was set at 0.05.

3. Results

3.1. Sample characteristics

Socio-demographic and clinical characteristics of the included and excluded patients are presented in Table 1. The excluded patients differed significantly from the included patients concerning diagnostic categories ($p < 0.001$), age at onset ($p = 0.012$) and illness duration ($p < 0.001$). The median DUP of the included patients was less than one month (range < 1 – 226; interquartile range 2). DUP had a heavily skewed distribution, with a majority of patients (63.1%) having a DUP of less than one month. Figure 1 illustrates the distribution of DUP.

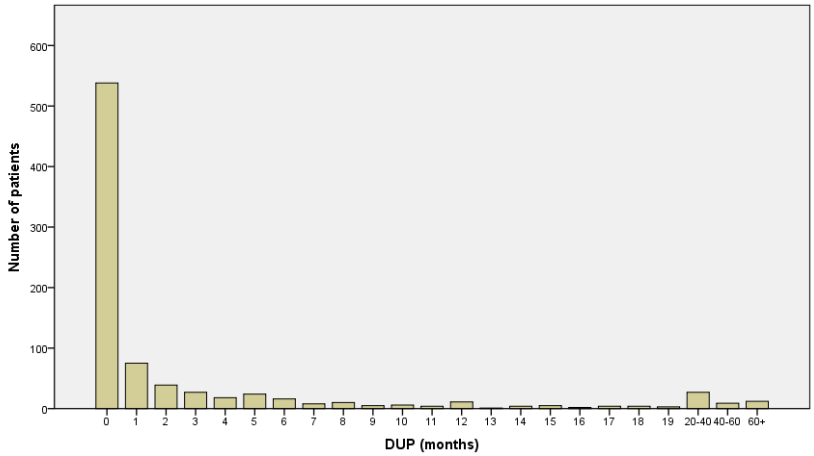


Figure 1: Distribution of the duration of untreated psychosis (N = 852).

Table 1: Socio-demographic and clinical characteristics of the included and excluded patients.

Variables	Included (N = 852)	Excluded (N = 258)	P-value
Age (years), mean (SD)	27.0 (7.2)	27.6 (7.6)	0.370[a]
Gender			0.172[b]
Male, N (%)	664 (77.9)	183 (73.8)	
Female, N (%)	188 (22.1)	65 (26.2)	
Educational level			0.388[c]
No education, N (%)	6 (0.7)	1 (0.4)	
Primary education, N (%)	106 (12.5)	37 (16.2)	
Lower vocational education, N (%)	95 (11.2)	31 (13.5)	
Lower general secondary education, N (%)	167 (19.7)	44 (19.2)	
Higher general secondary education, N (%)	114 (13.5)	18 (7.9)	
Pre-university secondary education, N (%)	107 (12.6)	30 (13.1)	
Intermediate vocational education, N (%)	142 (16.8)	38 (16.6)	
Higher vocational education, N (%)	74 (8.7)	23 (10.0)	
University, N (%)	36 (4.3)	7 (3.1)	
Migration status			0.286[b]
Native, N (%)	516 (76.0)	137 (70.3)	
Immigrant: first generation, N (%)	63 (9.3)	22 (11.3)	
Immigrant: second generation, N (%)	100 (14.7)	36 (18.5)	
Premorbid adjustment, mean (SD)	1.94 (0.90)	1.81 (0.87)	0.162[a]
Diagnostic categories			<0.001[c*]
Schizophrenia, N (%)	570 (67.1)	152 (63.1)	
Schizophreniform disorder, N (%)	50 (5.9)	12 (5.0)	
Schizoaffective disorder, N (%)	99 (11.6)	20 (8.3)	
Delusional disorder, N (%)	18 (2.1)	3 (1.2)	
Brief psychotic disorder, N (%)	19 (2.2)	10 (4.1)	
Substances induced psychosis, N (%)	4 (0.5)	1 (0.4)	
Psychosis NOS, N (%)	90 (10.6)	24 (10.0)	
Other, N (%)	0 (0.0)	19 (7.9)	
Age at onset (years), mean (SD)	22.7 (6.6)	21.3 (6.8)	0.012[a*]
Illness duration (years), mean (SD)	3.8 (3.4)	6.3 (5.1)	<0.001[a*]
DUP (months), median (range, IQR)	< 1 (< 1 – 226, 2)	…	…
Subdivision of DUP		…	…
Less than one month, N (%)	538 (63.1)		
One to three months, N (%)	141 (16.5)		
Three to six months, N (%)	58 (6.8)		
Six to twelve months, N (%)	44 (5.2)		
Twelve months and over, N (%)	71 (8.3)		

N = number; SD = standard deviation; NOS = not otherwise specified; IQR = interquartile range; DUP = duration of untreated psychosis; [a] = Mann Whitney test; [b] = Chi-square test; [c] = Fisher's exact test; *p < 0.05.

3.2. Factors contributing to DUP

The final logistic model produced by backward elimination included the variables gender, migration status and age at onset of the psychotic disorder. The score test for the proportional odds assumption of the final model was not significant (p = 0.112), meaning that the assumption was not violated.

As shown in Table 2, gender was significantly associated with DUP (p = 0.034). Since male was the reference category and the odds ratio was less than one, being male was associated with a longer DUP (odds ratio 0.65; 95% C.I. 0.44-0.97). Furthermore, being a first generation immigrant patient - compared to being a native patient - was associated with a longer DUP (p = 0.028; odds ratio 1.74; 95% C.I. 1.06-2.87). Being a second generation immigrant patient was not significantly associated with DUP (p = 0.868). Age at onset of the psychotic disorder was also significantly associated with DUP (p = 0.003). This represents that being younger at onset was associated with a longer DUP (odds ratio 0.96; C.I. 0.94-0.99).

Table 2: Factors within the final logistic model and their association with the duration of untreated psychosis.

Parameter	Estimate	S.E.	Wald X^2	P-value	Odds Ratio	95% C.I.
Intercept (DUP > 12)	-1.57	0.32	24.75	<0.001	…	…
Intercept (DUP 6 to 12)	-1.01	0.30	11.00	<0.001	…	…
Intercept (DUP 3 to 6)	-0.46	0.30	2.39	0.122	…	…
Intercept (DUP 1 to 3)	0.35	0.30	1.41	0.235	…	…
Gender						
(Female vs. male)	-0.43	0.20	4.49	0.034*	0.65	0.44 - 0.97
Migration status 1	0.56	0.25	4.81	0.028*	1.74	1.06 - 2.87
(First generation vs. native)						
Migration status 2	-0.04	0.22	0.03	0.868	0.96	0.62 - 1.49
(Second generation vs. native)						
Age at onset	-0.04	0.01	8.92	0.003*	0.96	0.94 - 0.99

S.E. = standard error; C.I. = confidence interval; DUP = duration of untreated psychosis; Migration status 1 = immigrant: first generation; Migration status 2 = immigrant: second generation; *p < 0.05.

4. Discussion

4.1. Main findings

DUP was longer for male patients, first generation immigrant patients and patients being younger at onset of the psychotic disorder. Educational level, premorbid adjustment and being a second generation immigrant were not significantly associated with DUP in our sample.

4.2. Comparison with previous studies

The median DUP of less than one month found in this study is short compared with the average median DUP of 26 weeks reported in most international studies (Marshall et al., 2005), but is in line with recently published Dutch studies (Wunderink et al., 2006; Boonstra et al., 2012b). There are several possible explanations for this difference with international studies. First this may partly be the merit of the Dutch and Belgium health care system, which is characterized by an easy service access. Health insurance is compulsory and there is an extensive system of hospital and community based mental health services, including in- and outpatient clinics, outreaching teams and psychiatric

emergency services. Second, DUP estimation varies greatly according to commonly used definitions (Polari et al., 2011). The most restrictive criteria for the start-point for DUP and the least restrictive criteria for the end-point for DUP, will lead to the shortest DUP estimation. In this study the start-point for DUP was defined by the onset of the first psychotic episode. The other common start-point used is the onset of the first psychotic symptoms. Since it is not uncommon that the first episode arises months to even years after the onset of the first psychotic symptoms, one can imagine that by choosing our definition the start-point moves towards the end-point, and DUP estimations will shorten. The end-point for DUP was defined as the initiation of antipsychotic medication and/or regular treatment contact with a mental health professional for psychosis. This definition is less restricted than many other definitions for the end-point for DUP, like for example "initiation of antipsychotic medication with good compliance". By choosing our definition the end-point for DUP moves towards its start-point, and DUP estimations will shorten. Furthermore, noting DUP in months instead of weeks or days, restricts its distribution even more. It would be more detailed to do otherwise, but we were restricted by the available data. Because there are no agreed-on cutoff points for DUP being (too) long (Marshall et al., 2005), it is unclear to what duration DUP must be reduced at least, in order not to increase the risk of a worse prognosis. In our next study we will analyze whether there is an association between this relatively short DUP and various outcome variables.

The finding that gender was associated with DUP was not expected given the findings of a review article describing the literature up to 2010 (Cascio et al., 2012), but is in line with more recent studies showing men to have (a tendency for) a longer DUP compared with women (Chang et al., 2011; Fridgen et al., 2012), and with other studies showing delayed help seeking in men (Galdas et al., 2004). It is conceivable that in men a lower level of awareness and insight to illness might contribute to their treatment delay (Cotton et al., 2006).

The finding that first generation immigrant patients had a longer DUP than native-born patients is in line with findings of three smaller studies (Sterk et al., 2010; Boonstra et al., 2012b; Nerhus et al., 2013). There are several explanations for this association. First generation immigrants may be less familiar with the concept of mental illness and the mental health services (Wolff et al., 1996a), may be less likely to perceive themselves as having a psychiatric problem or to be in need for treatment (Commander et al., 1999) and may experience more pronounced negative emotions towards mental illnesses and visiting a psychiatrist (Wolff et al., 1996b; Sadeghieh Ahari et al., 2013). Language barrier may also cause delay. Although all patients spoke sufficient Dutch to enter this study, bilingual individuals are often less able to express themselves in their second language when they are acutely psychotic (Paradis, 2008), which may impede diagnostics and treatment. One may argue that these factors diminish in following generations (Boonstra et al., 2012b), which may explain why second generation immigrant patients did not have a longer DUP compared with native-born patients. A recent meta-analysis showed little evidence to support an association between ethnicity and DUP (Anderson et al., 2013). The difference between the present study and this meta-analysis is that we focused on migration status, not on ethnic origin. Differences in DUP might therefore be related to the effects of migration, and not to ethnic background.

The association between age at onset of the psychotic disorder and DUP was expected, and is in line with previous findings (Bechard-Evans et al., 2007; Schimmelmann et al., 2008). However, it is important to note that the odds ratio was nearly one, and therefore age at onset may not be a clinically relevant risk factor of DUP.

The finding that educational level was not associated with DUP was expected, and is in line with most observations made by others (Morgan et al., 2006; Bechard-Evans et al., 2007; Compton et al., 2008; Compton et al., 2011).

The lack of association between premorbid adjustment and DUP was unexpected, but is in line with previous studies with (mainly) non-affective psychotic diagnoses samples (Chen et al., 2005; Harris et al., 2005)

4.3. Additional statistical analyses

An ordinal logistic regression analysis was applied to investigate factors contributing to DUP, but also several alternative analyses were applied.

DUP can be viewed as a time-to-event outcome, and it can therefore be analyzed with a Cox proportional hazards model. This analysis confirmed that first generation migration status and age at onset of the psychotic disorder are significantly associated with DUP, but did not confirm the association between gender and DUP. However, these results may be biased. Our DUP has a very high number of ties, which may affect the parameter estimates. Although the Efron's approximation method we used is considered the best possible approach for handling substantial numbers of ties (Hertz-Picciotto and Rockhill, 1997), this method may still result in serious biases (Scheike and Sun, 2007; Allison, 2010). It is therefore more difficult to interpret the results of this analysis compared to the results of ordinal logistic regression.

An alternative method is to treat DUP as a numerical outcome. Due to the high number of DUP values of less than one month, a zero-inflated regression model is suitable. Therefore a zero-inflated negative binomial model was applied. This analysis is more complex compared to ordinal logistic regression, but it supported its findings: the factors gender, migration status and age at onset of the psychotic disorder, were significantly associated with DUP.

Choosing the appropriate statistical analysis for DUP is not straightforward, since all analyses have their advantages and disadvantages. The ordinal logistic regression analysis was chosen because it has the easiest interpretation.

Several sensitivity analyses were conducted to determine the robustness of our conclusions. One sensitivity analysis investigated the effect of an additional category of "negative DUP". This analysis confirmed that first generation migration status and age at onset of the psychotic disorder are significantly associated with DUP, but did not confirm the association between gender and DUP. However, the assumption of proportionality was violated. This makes it difficult to value the results of this analysis. Another sensitivity analysis included patients with a calculated DUP exceeding the duration of the first psychotic episode. This analysis did not alter our conclusions.

4.4. Strengths and limitations

A major strength of this study is the large sample size and its representativeness for the treated prevalence of non-affective psychotic disorders. Concerning generalizability however, it is important to keep in mind that the included patients differed significantly from the excluded patients concerning age at onset and illness duration. Other strengths are the standardized instruments that were used and the use of a statistical model that takes into account the effect of other candidate factors. Using a categorical approach allowed inclusion of a number of patients with a very long DUP, which tend to get excluded in other studies. A limitation in all studies in which DUP is assessed, is that DUP is defined retrospectively and that data collection relies on self-reports. Also, we could not evaluate the contribution of "type of onset of the psychotic disorder" – a factor reported to be associated with DUP (Morgan et al., 2006; Compton et al., 2008). Furthermore, it would be of interest to explore whether ethnicity plays part in the association between DUP and migration status. This exploration could not be done within this study, because small numbers of the different ethnic groups precluded separate analyses.

4.5 Implications

Identifying factors contributing to a delayed identification of patients and a delayed start of treatment may be important, because it can give direction to early detection and early intervention initiatives. Although research into early detection programs is still sparse, the TIPS (Early Treatment and Intervention in First Episode Psychosis) researchers from Norway have demonstrated that reducing DUP is possible and that this might produce favorable long term improvements (Melle et al., 2004; Hegelstad et al., 2012).

The findings presented in this paper suggest that the process involved in men's help seeking behaviour deserves attention. Furthermore, the association we found between migration status and DUP may be of major importance in countries like the Netherlands, where immigrants constitute more than one fifth of the population (Centraal Bureau voor de Statistiek, 2013). Future research should aim to explore the explanations for this association and how to design interventions to shorten treatment delay.

4.6. Conclusions

In this study, it was shown that DUP was longer for patients being younger at onset of the psychotic disorder, for first generation immigrant patients and probably also for male patients. Further research is warranted to detect what explains these associations and what interventions are needed to shorten DUP and thereby possibly improve prognosis. In our large non-affective psychotic sample, educational level and premorbid adjustment were not significantly associated with DUP.

References

Allison, P.D., 2010. Survival analysis using SAS: a practical guide. 2nd Edition, Cary, NC: SAS Institute, USA

Anderson, K.K., Flora, N., Archie, S., Morgan, C., McKenzie, K., 2013. Race, ethnicity, and the duration of untreated psychosis: a systematic review. Soc. Psychiatry Psychiatr. Epidemiol. [Epub ahead of print]

Andreasen, N.C., Flaum, M., Arndt, S., 1992. The Comprehensive Assessment of Symptoms and History (CASH). An instrument for assessing diagnosis and psychopathology. Arch. Gen. Psychiatry 49 (8), 615-623.

Bechard-Evans, L., Schmitz, N., Abadi, S., Joober, R., King, S., Malla, A., 2007. Determinants of help-seeking and system related components of delay in the treatment of first-episode psychosis. Schizophr. Res. 96 (1-3), 206-214.

Boonstra, N., Klaassen, R., Sytema, S., Marshall, M., De Haan, L., Wunderink, L., Wiersma, D., 2012a. Duration of untreated psychosis and negative symptoms - a systematic review and meta-analysis of individual patient data. Schizophr. Res. 142 (1-3), 12-19.

Boonstra, N., Sterk, B., Wunderink, L., Sytema, S., De Haan, L., Wiersma, D., 2012b. Association of treatment delay, migration and urbanicity in psychosis. Eur. Psychiatry 27 (7), 500-505.

Bottlender, R., Sato, T., Jager, M., Wegener, U., Wittmann, J., Strauss, A., Moller, H.J., 2003. The impact of the duration of untreated psychosis prior to first psychiatric admission on the 15-year outcome in schizophrenia. Schizophr. Res. 62 (1-2), 37-44.

Cannon-Spoor, H.E., Potkin, S.G., Wyatt, R.J., 1982. Measurement of premorbid adjustment in chronic schizophrenia. Schizophr. Bull. 8 (3), 470-484.

Cascio, M.T., Cella, M., Preti, A., Meneghelli, A., Cocchi, A., 2012. Gender and duration of untreated psychosis: a systematic review and meta-analysis. Early Interv. Psychiatry 6 (2), 115-127.

Centraal Bureau voor de Statistiek, 2013. Bevolking; generatie, geslacht, leeftijd en herkomstgroepering. Homepage on the Internet. Updated 2013 May 21, last accessed 2013 December 18. Available from http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=37325&D1=a&D2=0&D3=0&D4=0&D5=04,137,152,220,237&D6=0,4,9,(l-1),l&HD=130605-0936&HDR=G2,G1,G3,T&STB=G4,G5

Chang, W.C., Tang, J.Y., Hui, C.L., Chiu, C.P., Lam, M.M., Wong, G.H., Chung, D.W., Law, C.W., Tso, S., Chan, K.P., Hung, S.F., Chen, E.Y., 2011. Gender differences in patients presenting with first-episode psychosis in Hong Kong: a three-year follow up study. Aust. N. Z. J. Psychiatry 45 (3), 199-205.

Chang, W.C., Tang, J.Y., Hui, C.L., Lam, M.M., Wong, G.H., Chan, S.K., Chiu, C.P., Chung, D.W., Law, C.W., Tso, S., Chan, K., Hung, S.F., Chen, E.Y., 2012. Duration of untreated psychosis: relationship with baseline characteristics and three-year outcome in first-episode psychosis. Psychiatry Res. 198 (3), 360-365.

Chen, E.Y., Dunn, E.L., Miao, M.Y., Yeung, W., Wong, C., Chan, W., Chen, R.Y., Chung, K., Tang, W., 2005. The impact of family experience on the duration of untreated psychosis (DUP) in Hong Kong. Soc. Psychiatry Psychiatr. Epidemiol. 40 (5), 350-356.

Chong, S.A., Mythily, Lum, A., Chan, Y.H., McGorry, P., 2005. Determinants of duration of untreated psychosis and the pathway to care in Singapore. Int. J. Soc. Psychiatry 51 (1), 55-62.

Commander, M.J., Cochrane, R., Sashidharan, S.P., Akilu F., Wildsmith, E., 1999. Mental health care for Asian, black and white patients with non-affective psychoses: pathways to the psychiatric hospital, in-patient and after care. Soc. Psychiatry Psychiatr. Epidemiol 34 (9), 484-491.

Compton, M.T., Chien, V.H., Leiner, A.S., Goulding, S.M., Weiss, P.S., 2008. Mode of onset of psychosis and family involvement in help-seeking as determinants of duration of untreated psychosis. Soc. Psychiatry Psychiatr. Epidemiol. 43 (12), 975-982.

Compton, M.T., Gordon, T.L., Goulding, S.M., Esterberg, M.L., Carter, T., Leiner, A.S., Weiss, P.S., Druss, B.G., Walker, E.F., Kaslow, N.J., 2011. Patient-level predictors and clinical correlates of duration of untreated psychosis among hospitalized first-episode patients. J. Clin. Psychiatry 72 (2), 225-232.

Corrigan, P., 2004. How stigma interferes with mental health care. Am. Psychol. 59 (7), 614-625.

Cotton, S.M., Wright, A., Harrig, M.G., Jorm, A.F., McGorry, P.D., 2006. Influence of gender on mental health literacy in young Australians. Aust. N. Z. J. Psychiatry 40 (9), 790-796.

Crumlish, N., Whitty, P., Clarke, M., Browne, S., Kamali, M., Gervin, M., McTique, O., Kinsella, A., Waddington, J.L., Larkin, C., O'Callaghan, E., 2009. Beyond the critical period: longitudinal study of 8-year outcome in first-episode non-affective psychosis. Br. J. Psychiatry 194 (1), 18-24.

Drake, R.J., Haley, C.J., Akhtar, S., Lewis, S.W., 2000. Causes and consequences of duration of untreated psychosis in schizophrenia. Br. J. Psychiatry 177 (6), 511-515.

Fridgen, G.J., Aston, J., Gschwandtner, U., Pflueger, M., Zimmermann, R., Studerus, E., Stieglitz, R.D., Riecher-Rössler, A., 2012. Help-seeking and pathways to care in the early stages of psychosis. Soc. Psychiatry Psychiatr. Epidemiol. Article not assigned to an issue. Published online 2012 December 25, last accessed 2013 December 18. DOI 10.1007/s00127-012-0628-0.

Galdas, P.M., Cheater, F., Marshall, P., 2004. Men and health help-seeking behaviour: literature review. J. Adv. Nurs. 49 (6), 616-623.

Harrigan, S.M., McGorry, P.D., Krstev H., 2003. Does treatment delay in first-episode psychosis really matter? Psychol. Med. 33 (1), 97-110.

Harris, M.G., Henry, L.P., Harrigan, S.M., Purcell, R., Schwartz, O.S., Farrelly, S.E., Prosser, A.L., Jackson, H.J., McGorry, P.D., 2005. The relationship between duration of untreated psychosis and outcome: An eight-year prospective study. Schizophr. Res. 79 (1), 85-93.

Hegelstad, W.T., Larsen, T.K., Auestad, B., Evensen, J., Haahr, U., Joa, I., Johannesen, J.O., Langeveld, J., Melle, I., Opjordsmoen, S., Rossberg, J.I., Rund, B.R., Simonsen, E., Sundet, K., Vaglum, P., Friis, S., McGlashan, T., 2012. Long-term follow-up of the TIPS early detection in psychosis study: effects on 10-year outcome. Am. J. Psychiatry 169 (4), 374-380.

Hertz-Picciotto, I., Rockhill, B., 1997. Validity and efficiency of approximation methods for tied survival times in Cox regression. Biometrics 53 (3), 1151-1156.

Korver, N., Quee, P.J., Boos, H.B., Simons, C.J., De Haan, L., GROUP investigators, 2012. Genetic Risk and Outcome of Psychosis (GROUP), a multi-site longitudinal cohort study focused on gene-environment interaction: objectives, sample characteristics, recruitment and assessment methods. Int. J. Methods Psychiatr. Res. 21 (3), 205-221.

Marshall, M., Lewis, S., Lockwood, A., Drake, R., Jones, P., Croudace, T., 2005. Association between duration of untreated psychosis and outcome in cohorts of first-episode patients: a systematic review. Arch. Gen. Psychiatry 62 (9), 975-983.

Melle, I., Larsen, T.K., Haahr, U., Friis, S., Johannessen, J.O., Opjordsmoen, S., Simonsen, E., Rund, B.R., Vaglum, P., McGlashan, T., 2004. Reducing the duration of untreated first-episode psychosis: effects on clinical presentation. Arch. Gen. Psychiatry 61 (2), 143-150.

Morgan, C., Abdul-Al, R., Lappin, J.M., Jones, P., Fearon, P., Leese, M., Croudace, T., Morgan, K., Dazzan, P., Craig, T., Leff, J., Murray, R., AESOP Study Group, 2006. Clinical and social determinants of duration of untreated psychosis in the AESOP first-episode psychosis study. Br. J. Psychiatry 189 (5), 446-452.

Nerhus, M., Berg, A.O., Haram, M., Kvitland, L.R., Andreassen, O.A., Melle, I., 2013. Migrant background and ethnic minority status as predictors for duration of untreated psychosis. Early Interv. Psychiatry. [Epub ahead of print]

Paradis, M., 2008. Biligualism and neuropsychiatric disorders. J. Neurolinguistics 21 (3), 199-230.

Perkins, D.O., Gu, H., Boteva, K., Lieberman, J.A., 2005. Relationship between duration of untreated psychosis and outcome in first-episode schizophrenia: a critical review and meta-analysis. Am. J. Psychiatry 162 (10), 1785-1804.

Polari, A., Lavoie, S., Sarrasin, P., Pellanda, V., Cotton, S., Conus, P., 2011. Duration of untreated psychosis: a proposition regarding treatment definition. Early Interv. Psychiatry 5 (4), 301-308.

Sadeghieh Ahari, S., Nikpou, H., Molavi, P., Abdi, R., Amani, F., Shirinzadeh, B., 2014. An investigation of duration of untreated psychosis and the affecting factors. J. Psychiatr. Ment. Health Nurs. 21 (1), 87-92.

Sartorius, N., Gulbinat, W., Harrison, G., Laska, E., Siegel, C., 1996. Long-term follow-up of schizophrenia in 16 countries. A description of the International Study of Schizophrenia conducted by the World Health Organization. Soc. Psychiatry Psychiatr. Epidemiol. 31 (5), 249-258.

Scheike, T.H., Sun, Y., 2007. Maximum likelihood estimation for tied survival data under Cox regression model via EM-algorithm. Lifetime Data Anal. 13 (3), 399-420.

Schimmelmann, B.G., Huber, C.G., Lambert, M., Cotton, S., McGorry, P.D., Conus P., 2008. Impact of duration of untreated psychosis on pre-treatment, baseline, and outcome characteristics in an epidemiological first-episode psychosis cohort. J. Psychiatr. Res. 42 (12), 982-990.

Sterk, B., Slief, E.M., Blankers, M., Linszen, D.H., De Haan, L., 2010. Duration of untreated psychosis and ethnicity. Schizophr. Res. 124 (1-3), 238-239.

Tanskanen, S., Morant, N., Hinton, M., Lloyd-Evans, B., Crosby, M., Killaspy, H., Raine, R., Pilling, S., Johnson, S., 2011. Service user and carer experiences of seeking help for a first episode of psychosis: a UK qualitative study. B.M.C. Psychiatry 11, 157-167.

Verdoux, H., Bergey, C., Assens, F., Abalan, F., Gonzales, B., Pauillac, P., Fournet, O., Liraud, F., Beaussier, J., Gaussares, C., Etchegaray, B., Bourgeois, M., Van Os, J., 1998. Prediction of duration of psychosis before first admission. Eur. Psychiatry 13 (7), 346-352.

Verhage, F., 1964. Intelligentie en leeftijd: Onderzoek bij Nederlanders van twaalf tot zevenenzeventig jaar. Van Gorcum, Assen.

Wing, J.K., Babor, T., Brugha, T., Burke, J., Cooper, J.E., Giel, R., Jablenski, A., Regier, D., Sartorius, N., 1990. SCAN. Schedules for Clinical Assessment in Neuropsychiatry. Arch. Gen. Psychiatry 47 (6), 589-593.

Wolff, G., Pathare, S., Craig, T., Leff, J., 1996a. Community knowledge of mental illness and reaction to mentally ill people. Br. J. Psychiatry 168 (2), 191-198.

Wolff, G., Pathare, S., Craig, T., Leff, J., 1996b. Community attitudes to mental illness. Br. J. Psychiatry 168 (2), 183-190.

Wunderink, A., Nienhuis, F.J., Sytema, S., Wiersma, D., 2006. Treatment delay and response rate in first episode psychosis. Acta. Psychiatr. Scand. 113 (4), 332-339.

CHAPTER 8

Summary and General Discussion

Aims of the thesis

The clinical presentation of psychotic disorders (e.g. on positive, negative symptoms or cognitive impairment) is heterogeneous and symptoms differ in origin, structure and in clinical expression. Although this notion has been widely acknowledged (Markova and Berrios, 1995), symptom differences are still overlooked in clinical practice and proper statistical analysis in research is lacking. The heterogeneity of psychotic disorders calls for adequate statistical approaches to clarify the underlying structures. Clustering techniques or group-based trajectory techniques have been implemented to form meaningful homogeneous symptom subtypes of subjects using longitudinal data (Goldstein and Shemansky, 1995; Jablensky, 2006; Joyce and Roiser, 2007; Dawes et al., 2011; Quee et al., 2014). The main aim of the thesis is to explore heterogeneity in cognitive functioning and in clinical symptoms in schizophrenia patients and their unaffected siblings using both cross-sectional *and* longitudinal data.

To reach the objective of the thesis, a number of different steps were taken. Firstly, I compared the performance of 14 cluster indices to identify the right number of cognitive subtypes. For this, I performed simulations of various number of clustering scenarios, based on a real case study with cognitive measures (**Chapter 2**). Secondly, I explored the heterogeneity of clinical symptoms and of cognitive functioning in patients with psychosis and their unaffected siblings in a longitudinal setting by implementing the group-based trajectory modeling approach (**Chapter 3** and **4**). Thirdly, I examined also the unobserved heterogeneity of symptoms by using factors (neuro and social cognition) to predict the development of psychotic experiences over time by applying mixture of generalized linear mixed effects modeling (**Chapter 5**). Fourthly, I described the heterogeneity in the domain of comorbid diseases among patients with schizophrenia, their unaffected siblings, and healthy controls. The effects of gender, age and familial liability on the prevalence of multimorbidity were also investigated (**Chapter 6**). Finally, I studied the risk factors that were associated with the duration of untreated psychosis (DUP) in a large sample that represented the treated prevalence of non-affective psychotic disorders (**Chapter 7**). All studies in this thesis were performed within the framework of the Genetic Risk and Outcome of Psychosis (GROUP) project, a longitudinal multicenter cohort study in the Netherlands and Belgium.

Summary of main findings

The first major challenge was to correctly identify the number of clusters in a complex heterogeneous dataset. Over the last decades, different indices have been developed to quantify the number of clusters. In **chapter 2**, I investigated the most promising indices for detecting the correct number of clusters on the basis of hierarchical clustering (with Ward's agglomerative method) in cross-sectional data. The indices were investigated on *(i)* how well they would discriminate between a single and multiple cluster solution and *(ii)* on how well they can predict the number of clusters in a multi cluster solution. I showed that, out of fourteen indices, the Duda and Hart (DH), Hartigan (H), and Gap/pc indices were best performing in the simulation study involving eight-dimensional cognitive outcome variables taken from a real case study of schizophrenic patients. These indices predicted with high probabilities the simulated number of clusters within the range of one cluster difference from the real case (actual number of clusters). The DH index was the most consistent, while Gap/pc

in combination with WGap/pc was capable of answering the question if a multiple cluster solution was present or not.

Next, I derived subtypes of patients with schizophrenia and their unaffected siblings longitudinally in terms of cognitive functioning (composite score) and in terms of clinical symptoms. For this, in **chapter 3,** I applied group-based trajectory modeling. Five trajectories of patients were found. These trajectories were labeled as 'normal' (26.7%) as their z-scores for the composite cognitive measures were in the standard normal range; 'mild alterations' (30.4%) as their performances were 0.5 SD below normal; 'moderate impairment' (28.4%) as their z-scores ranged 1 SD below normal; 'severe impairment' (10.7%) as their z-scores were more than 1 SD below normal. A small group of patients was labeled 'high performer' (3.8%), as they performed better than healthy controls. In a similar vein, four trajectories for siblings were identified. They were labeled as 'normal' (37.6%), 'mild alterations' (25.1%), 'moderate impairment' (13.0%) and 'high performer' (24.2%). These distinct trajectories of both patients and siblings turned out to be stable and persistent over time. Impaired patients and siblings were from ethnic minorities, younger age, low IQ and exhibited poorer premorbid functioning compared to the normal group. Severely impaired patients with schizophrenia also had more severe symptomatology, (*i.e.* poorer performance on PANSS five-factor model). Next, I hypothesized that subtype of patient predicted the subtype of sibling within the sibling-patient analysis. I considered sibling subtype (multi-category) as dependent and patient subtype as independent categorical variables. Given the family structure of the data (as siblings-patients belong within the same family), clustered multinomial logistic regression analysis was conducted taking into account family membership as a random effect. Results showed the familial correlation (*i.e.* the intra-class correlation coefficient between pairs of index patients and their unaffected siblings) accounted for 27 percent of total variation. I showed that cognitive subtypes of patients significantly predicted the cognitive subtypes of siblings. The poorer the cognitive profile of a patient, the better it predicted (OR 10.07, 95% CI 4.15-24.44) that of a more cognitively impaired sibling. Similarly, patients with moderately impaired cognition predicted his/her unaffected siblings to be moderately cognitive impaired (OR 5.7, 95% CI 2.77-11.70). Another finding was that severely impaired patients predicted much less (OR 0.24, 95% CI 0.09-0.63) the sibling to be a higher performer. The relative risks of mild alterations and moderate impaired group of patients also offered less predictive value to the high performance profiles of siblings. It is conceivable that the high performance of unaffected siblings is unlikely to be predicted given the status of their paired probands.

Besides investigating the cognitive heterogeneity, I also further explored the heterogeneity in negative symptoms. Recent literature has revealed that in fact there are two subdomains of negative symptoms, typically known as social amotivation (SA) and expressive deficits (ED) (Messinger et al., 2011; Foussias et al., 2014; Liemburg et al., 2013). To date, there are few studies describing the longitudinal course of SA and ED, and conclusions are mixed (Ergul and UCok, 2015; Norman et al., 2015; Galderisi et al., 2013). In **chapter 4**, I investigated *i)* whether the course of SA and ED changed during the course of six years. Using linear mixed models, I found that both SA and ED displayed a small but significant decrease in severity over time. Next, I examined *ii)* whether baseline levels of SA and ED could predict functioning and quality of life six years later. Multiple linear regression analysis

was conducted and the results indicated that lower baseline levels of SA symptoms predicted higher levels of global functioning, of social functioning, better quality of life and more engagement in work or study activity six years later. In the same line, lower baseline levels of ED symptoms predicted higher levels of global and social functioning six years later.

I then aimed *iii)* to identify subgroups based on the course of SA and ED over a period of six years. Here, I again applied group-based trajectory modeling in order to identify the correct number of subgroups. Each of the two subdomains yielded four subgroups with differing courses of negative symptom levels. Within both subdomains, a large group of patients (±60%) with steady low levels of symptoms and two subgroups (both ±15%) with either symptom-level increase or symptom-level decrease were found. Furthermore, within SA, a small subgroup (±6%) showed decreasing symptom levels over time after having started at a higher symptom levels, whereas within the ED subdomains, another small group (±6%) was continuously experiencing high symptom levels.

I further investigated *iv)* the relationship between subgroups within SA and ED and functioning and quality of life over the course of six years. Linear and generalized linear mixed effects modeling were conducted to unravel the relationships. The stable-low SA group had better functional outcome as measured by the Global Assessment of Functioning (GAF), Social Functioning Scale (SFS) and World Health Organization Quality of Life (WHO-QOL) at all time points. The stable-low ED patient-group also performed way better on the GAF and SFS, at baseline and after three years of follow up, but not six-years after baseline. People with decreased-to-low SA-symptoms had a significantly lower chance of living independently after three years of follow up and reported higher levels of both, global functioning and quality of life after six years compared to low SA symptoms. People with an increased SA had lower chance to have a regular work activity and reported lower functioning and lower quality of life over the course of six years. In a similar way, patients with increased ED showed significantly lower chance of work activity and lower levels of functioning than low ED over time. In addition, decreased ED was associated with better global functioning over the course of six years.

In summary, results show that there is considerable heterogeneity in the course of subdomains over time and suggest that negative symptoms are less stable than was previously assumed. Subgroups were identified within SA and ED, showing a different course of symptoms over time. Moreover, the two domains are clinically relevant as they differentially relate to the level and course of outcomes.

Besides cognitive and symptoms heterogeneity, I also studied psychotic experiences in relatives of individuals with psychotic disorders. Psychotic experiences are heterogeneous over time and this heterogeneity may be explained by different factors (e.g. neuro- and social cognition). To learn more about the development of psychotic experiences over three years, I investigated factors that were known to predict psychosis. To this end, in **chapter 5**, I applied mixture of generalized linear mixed effects models to examine the reported development of psychotic experiences in siblings of people with psychosis and its relationships to neuro- and social cognition. Poorer verbal learning performance (operationalized as the score on immediate recall) predicted the occurrence of PE after three years and also the distress associated with these psychotic experiences. Moreover, better baseline performance on a hinting task, representing Theory of Mind (ToM) was associated

with a decrease in psychotic experiences three years later. Baseline distress was associated with poorer recognition of angry and neutral faces and strikingly, with better performance on the Benton facial recognition test at baseline. In conclusion, verbal learning and ToM were found to be predictive of frequency and course of psychotic experiences over three years respectively. In addition, verbal learning and ToM were also predictive of the distress that psychotic symptoms were causing.

In **chapter 6,** the heterogeneity in the domain of somatic diseases and complaints among patients with schizophrenia, their unaffected siblings and subjects from the healthy population is described. I investigated the effect of familial liability to psychosis along with gender and age on the prevalence of multimorbidity of diseases and complaints. In summary, familial liability had a significant effect on multimorbidity of complaints and lifetime diseases (with/without psychiatric comorbidity) respectively. Moreover, multimorbidity was strongly associated with female gender overall. In people with psychosis, multimorbidity was not limited to the elderly but also affected young individuals. The risk of multimorbidity caused by familial liability for psychosis was consistent across gender and age group, meaning that familial liability for psychosis is in itself a strong independent determinant of multimorbidity.

Finally, in **chapter 7,** I identified risk factors that were significantly associated with the duration of untreated psychosis (DUP). DUP is defined as the time from the emergence of the first psychotic episode to the initiation of adequate treatment. Migration status, age at onset of psychotic disorder and gender were significantly associated with DUP. In conclusion, first generation immigrant patients, individuals with an early onset of their psychotic disorder and male patients were at risk of a longer DUP. Our findings were in line with the results of a number of smaller studies (Sterk et al., 2010; Boonstra et al., 2012; Nerhus et al., 2015; Schimmelmann et al., 2008; Cotton et al., 2006). First generation immigrants seem to be less familiar with the concept of mental illness and also with mental health services (Wolff et al., 1996). They may be less likely to perceive themselves as having a psychiatric problem or to be in need for treatment (Commander et al., 1999).

From heterogeneity to endophenotype

Cognitive impairment and negative symptoms are core features of the presence and severity of psychosis (American Psychiatric Association, 2013). There is ample evidence for significant cognitive heterogeneity in schizophrenia. This heterogeneity is explained by a general loss of function, varying from patient to patient, or by impairment over the various cognitive abilities (*e.g.* executive function and working memory) (Joyce et al., 2005; Joyce and Roiser, 2007). Previous literature reported that the cognitive test performance of patients with schizophrenia was extremely heterogeneous (Goldstein, 1990; Joyce et al., 2005; Joyce and Roiser, 2007). So far, only two studies demonstrated cognitive heterogeneity in the *relatives* of patients with schizophrenia (Quee et al., 2014; Sautter et al., 1995). I extended the findings of Quee *et al*. (2014) by showing that cognitive heterogeneity in people with schizophrenia and in their unaffected siblings was stable for the long-term course.

Another type of heterogeneity (*i.e.* negative symptom heterogeneity) has also been observed in psychosis (Seaton et al., 1999; Seaton et al., 2001; Liemburg et al., 2013; Stiekema et al., 2016). In both cases, I confirmed the stable groupings of patients. The heterogeneity was confirmed by examining the relationships with course of illness, clinical, and functional outcomes. Therefore, these groupings may be considered as candidate subtypes.

Furthermore, I underpinned the arguments for cognitive subtypes to be considered good "endophenotypes" by fulfilling the following criteria (Miller and Rockstroh, 2013; Ritsner and Gottesman, 2011; Hasler et al., 2006):

i) stable trait: cognition as well as cognitive subtypes were shown to be highly stable over time and associated with illness of schizophrenia, *ii*) familial liability: the cognitive performance of family members of patients with schizophrenia is located in between cognitive performances of patients and healthy population (i.e. familial correlation), *iii*) cognitive subtypes were correlated with other clinical and socio-demographic parameters, and *iv*) cognitive functioning cannot be seen by the "naked eye". Thus, unraveling the heterogeneity of cognition yielded meaningful cognitive subtypes for both patients and their unaffected siblings, so called endophenotypes, being located somewhere between the genotype (DNA) and the phenotype (clinical symptoms).

Regarding negative symptoms as well as symptom subtypes: they are less stable but still persistent over time. Their significant relationships with other external parameters define them as subtype. However, they cannot be considered as endophenotype of schizophrenia, as they are part of the phenotype.

Not one but many forms of schizophrenias

A number of analytical techniques have been used to address the issue of cognitive heterogeneity. One option is to categorize cognitive functioning into neuro-psychologically normal and impaired subgroups based on standard cut-off scores for cognitive batteries or expert's judgment of cognitive profile (Seaton et al., 2001). However, categorizing patients into normal or impaired groups supposedly underestimates the actual heterogeneity in schizophrenia based on expert's cut-off scores. This limitation is addressed adequately using exploratory or model-based statistical techniques. Therefore, an alternative and more objective option is the application of clustering, like the model-based trajectory techniques to classify homogeneous cognitive subtypes to identify disease severity (Goldstein and Shemansky, 1995; Seaton et al., 1999; Seaton et al., 2001; Jablensky, 2006; Joyce and Roiser, 2007; Dawes et al., 2011; Quee et al., 2014). A number of studies in literature used cluster analysis (*e.g*. hierarchical or K-means) while other used latent profile analysis and reported four or five-cluster subtypes of patients with schizophrenia. These findings were stable and consistent although different studies used different test batteries (Goldstein, 1990; Heinrichs and Awad, 1993; Heinrichs et al., 1997; Goldstein et al., 1998; Seaton et al., 2001; Joyce et al., 2005; Joyce and Roiser, 2007; Bora et al., 2016). Likewise, several researchers investigated at least four potential subtypes of clinical symptoms (*e.g*. on the PANSS) of schizophrenia such as positive, negative, mixed and disorganized symptoms (Dollfus et al., 1996; Lykouras et al., 2001; Seaton et al., 2001). Moreover, there is also evidence that the negative symptoms are heterogeneous, yielding two subdomains *i.e.* social amotivation (SA) and expressive deficit (ED) (Liemburg et al., 2013; Stiekema et al., 2016). By group-based trajectory modeling, I confirmed that indeed patients are heterogeneous in both subdomains supporting the two dimensional approach of Liemburg *et al.* However, the DSM-V (American Psychiatric Association, 2013) favours one dimension.

Together with neurocognition and clinical symptoms, psychotic experiences of siblings of patients with schizophrenia are heterogeneous and are explained by neuro- and social-cognition (Appels et al., 2003; Kremen et al., 1994; Meijer et al., 2012; Quee et al., 2014; Snitz et al., 2006). In

line with the findings of our study (**chapter 5**), one study found that poorer theory of mind predicted the experience of delusions in children with auditory hallucinations after three years (Bartels-Velthuis et al., 2011). Fragmenting frequency of psychotic experiences and distress according to their true distributions and their associations with neuro-and social cognition, I conclude that psychotic experiences are heterogeneous over time. Apart from heterogeneity of clinical symptoms, I conclude that domains of somatic diseases and complaints are heterogeneous according to familial liability group as the prevalence of multimorbidity of complaints and lifetime diseases in siblings fall in between the prevalences of patients with schizophrenia and member of healthy population. One study showed the familial liability (controls<siblings<patients) as a risk factor for cognitive functioning (Krabbendam et al., 2001), although this is not the same as my conclusions on multimorbidity.

Based on findings from chapter 2 to 5, I conclude in line with Bleuler's hypothesis (1950) that there are many "schizophrenias" rather than a single form of schizophrenia (Bleuler, 1950). There are several considerations that would favor the cognition or symptom subtypes in schizophrenia rather than a continuum form of schizophrenia. First of all, the cluster analysis and trajectory modeling showed both level and pattern differences *on external variables* such as age, education, IQ, age of onset and premorbid functioning. Secondly, although there are clear differences in level of performances, the differences in average standardized cognition scores between patients and siblings are often dramatic, making it hard to imagine that there would be one underlying continuum of cognitive performance. Therefore, current data are in favor of cognitive heterogeneity containing multiple subtypes.

Methodological considerations

The strengths and limitations of the studies in this thesis have been addressed chapter-wise. Overall, the clustering techniques are appropriate to identify subtypes. The methodological strength in chapter 2 is that of applying extensive simulation studies that were based on a case study of eight dimensional measures to determine the number of clusters, making this a "real world" exercise. So far, performances of indices were studied mainly with artificial low dimensional simulated data (Milligan and Cooper, 1985; Tibshirani et al., 2001). In general, I studied only the indices which would fit best with the relative simple but frequently used methods of clustering. Thus, the study investigated the proposed indices using a sequential stopping criterion for hierarchical clustering. This means cutting the branches in the dendrogram horizontally. However, cutting branches can also be performed more dynamically making cuts higher and lower in the tree and not at the same height for the whole tree. It would be challenging and highly interesting to find out whether dynamic cutting would improve the performance of certain indices. Moreover, I focused only on Euclidean distance measure and Ward's agglomerative technique for identifying the number of clusters. It may be possible that some of the indices will provide different results when other distances or dissimilarity measures are used to merge subgroups within hierarchical clustering.

The strength of group-based trajectory modeling (GBTM) (Nagin, 2014) for analyzing developmental cognitive trajectories over time is that it is easily understood graphically and it provides simple and straightforward tabular data summaries. This method does take into account the missing outcome, covariates and patterns within the same model (Haviland et al., 2011) whilst other techniques such as growth-curve modeling or latent profile analysis do not.

Some limitations should also be mentioned in this regard. Sometimes GBTM provides a very small group of individuals which will not be representative for further analysis. For example, I found five meaningful subtypes in patients according to logged Bayes factor, but a small group of patients (3.8%) with high cognitive performance did not predict moderate impairment of siblings using sib-pair analysis. This may imply that the high performers in patients are a cluster artifact, but this is difficult to establish in unsupervised clustering.

Composite scores were computed using mean z-scores for all eight cognitive measures that were age and gender specific. We did not adjust for education as we believe that years of education is a measure, albeit rough, of "prodromal" cognitive functioning in schizophrenia. For that reason, it would not make sense to artificially control for years of education (Seaton et al., 2001).

Generating composite cognition scores might have an impact on finding meaningful trajectories instead of using multivariate cognitive tests. This is one of the limitations of trajectory modeling (Jones et al., 2001) which may deal with only one variable at a time measured over at least three assessments. Thus, here information may have been lost by using a composite score.

There may have been some selection bias in data collection with respect to patients or siblings compared to controls, as most of the controls were selected by random mailing.

Several additional limitations should also be mentioned on the subtypes for negative symptoms that were presented in chapter 4. We do not know whether changes in negative symptoms are due to relief of secondary negative symptoms, for example by reduced positive symptoms, depressive symptoms or antipsychotic medication (Carpenter and Kirkpatrick, 2015). Further, the duration between assessments is three years, which is large. Short interval may provide a clearer picture of negative symptoms persistency instead of observing at three years interval.

More importantly, there is a distinction between clusters and subtypes. If one conducts cluster or trajectory analysis and clusters/trajectories are determined, it does not mean that they reflect actual subtypes because cluster membership may be determined largely by level of performance. An extensive external validity must be established when cluster/trajectory becomes a subtype. For example, in schizophrenia, we usually want to see clinical, cognitive, neurobiological, and genetic evidence of the stability of the cluster/trajectory under study.

The assessment of psychotic experience at three-year follow-up might have been biased, resulting in observed decreases in frequency and distress of psychotic experiences. There may also be reporting bias since psychotic experiences are very personal experiences and thoughts which were assessed with a self-reported questionnaire.

The overall strength of the study is the use of a broad range of neurocognitive variables together with social cognition of healthy siblings of patients with schizophrenia, providing a comprehensive picture of the possible factors related to psychotic experiences. The methodology of using a mixture of generalized linear mixed effects models is unique for the predictive values of neuro- and social cognitive parameters on psychotic experiences. Measuring psychotic experiences in siblings yielded lot of zeros (i.e. no experiences) together with psychotic experiences ranging from 'sometimes' to 'nearly always'. In this situation, the weighted scores of psychotic experiences displayed a highly heterogeneous distribution function. Dealing with the heterogeneous outcomes, a

mixture of generalized linear mixed effects models (Tooze et al., 2002) would provide unbiased results.

The multimorbidity study was mainly based on self-reported diseases, clinical complaints and medication history. There might be a reporting bias regarding the diagnosis of the disease. The operational definition of multimorbidity is another concern. The majority of studies have defined multimorbidity as two or more, whereas others counted three or more concurrent diseases (Fortin et al., 2005; Jacobi et al., 2004; Fuchs et al., 2012; Willadsen et al., 2016). In this thesis, multimorbidity is defined as two or more diseases or complaints. Another concern is that I considered only gender, age and familial liability, leaving out other reported multiple risk factors (Agborsangaya et al., 2012; De Hert et al., 2011) for multimorbidity.

One of the strengths of this study was statistical methodological point of view. I applied generalized linear mixed effects models taking into account family structure as random effect to identify the risk factors for multimorbidity in psychosis. Additionally, this is the first study which has dealt with a comprehensive list of diseases, either somatic or psychiatric, in schizophrenia patients while counting multimorbidity in a cumulative way rather than focusing only on pairwise comorbidities (Nuyen et al., 2006; Oreski et al., 2012). The most informative feature of this study was the sibling model; whereas most studies emphasized multimorbidity either in healthy subjects or the disease population.

A major strength of duration of untreated psychosis (DUP) is the use of a statistical model that takes into account the effect of other candidate factors. I applied ordinal logistic regression for categorical DUP and a sensitivity analysis using Cox-proportional hazards model for time-to-event DUP. Modeling the data with equally valuable tools and then showing consistent results, makes the conclusion on the significant association of first generation migration status and age at onset of the psychotic disorder with DUP stronger than just using one analysis technique. One of the concerns is that categorization of DUP; there are no agreed-on cutoff points (Marshall et al., 2005). Other concerns of the results are that DUP was defined retrospectively and that data collection relied on self-reports.

There is a potential loss of statistical power and efficiency when missing data are present. This may lead to biases and incorrect statistical inferences. In chapter 4 and chapter 5, there were missing data both on outcomes and independent variables. Both chapters are assumed that the missingness is conditional on the observed data but independent of the unobserved values, which is called missing at random (MAR) (Allison, 2002). Maximum likelihood (ML) and multiple imputation (MI) (Rubin, 1987) methods usually handle the missingness under MAR. I employed a fully conditional specification (FCS) predicted mean matching (PMM) multiple imputation (MI) method to impute missing values for both continuous and categorical variables (van Buuren, 2007) and analyzed the data with appropriate statistical models. Apart from PMM, Bayesian MI method is another good approach to handle missingness under MAR assumptions. It uses an iterative algorithm to impute data and it splits the multivariate missing problem into a series of univariate problems based on the assumed distribution of the multivariate missing variables (e.g. multivariate normal for continuous variables, multinomial loglinear for categorical variables) (Schafer, 1997). However, if the missingness is Missing Not At Random (MNAR), i.e. missingness depends on the unobserved data and missingness

is no longer ignorable, a model for the probability of missing data needs to be specified. This probability model is then combined with a linear mixed model for the measurement process. Selection and pattern mixture models are two alternative and important approaches for dealing with MNAR. A selection model is the joint distribution of the measurement and the missing mechanisms into the marginal measurement distribution and the dropout distribution conditional on the measurements (Thijs et al., 2002; Verbeke and Molenberghs, 2000). A pattern-mixture model is the joint distribution of the measurement and response mechanisms into a different measurement model for all response patterns, and the marginal response distribution (Thijs et al., 2002; Michiels et al., 2002; Verbeke and Molenberghs, 2000). Pattern mixture model is a sensitivity analysis within a fully Bayesian modeling framework that could be used to handle the missingness under MAR (Little, 1995; Little, 1993; Michiels et al., 2002; Thijs et al., 2002).

Which clustering technique to use?

There are several clustering algorithms and each of them uses different induction principles. Literatures suggest that clustering methods are categorized into hierarchical, partitioning methods, model-based and grid-based methods (Fraley and Raftery, 1998; Han and Kamber, 2001; Fraley and Raftery, 2002; Estivill-Castro and Yang, 2000; Rokach and Oded, 2005).

A strength of hierarchical cluster analysis is that it always confirms that the most similar observations are in the same clusters. However, this is also its major weakness. Once a cluster is made, the observations within this cluster will never be relocated to other clusters. This problem does not occur with partitioning methods *e.g.* K-means or K-medoids. Additionally, partitioning methods are easy to interpret, simply to implement and faster to compute with large datasets. However, this method is sensitive to noisy data and outliers. Model-based or density-based methods, *e.g.* latent class/profile analysis, trajectory and growth-curve modeling, assume that observations would come from a mixture of distributions (Muthen and Shedden, 1999; Muthen and Muthen, 2000; Rokach and Oded, 2005). Latent class/profile analysis, trajectory or growth-curve modeling are subject-specific finite mixture modeling approaches, meaning that the focus is on the subject's unique pattern of characteristics and therefore focuses largely on identifying subtypes of individuals with similar patterns (Muthen and Muthen, 2000; Rokach and Oded, 2005; Jung and Wickrama, 2008). The model-based approaches assume a certain type of mixtures of distributions, while partitioning methods do not make any assumption on the distribution. Thus, the model-based approaches are theoretically superior to any other approaches when the distributional assumption is known. The algorithm takes into account the uncertainty when allocating observations to clusters. These approaches use to estimate the posterior probabilities that an individual will be categorized into a particular group of risk patterns (Shah et al., 2014). However, all approaches need a pre-defined number of clusters. Hierarchical clustering can be considered a prior analysis of K-means or latent profile analysis in cross-sectional studies and this thesis showed that it may be suitable to detect the number of clusters. Other clustering approaches e.g. decision trees, neural networks and grid-based methods are beyond the scope of this thesis.

Clinical implications

Cognitive functioning is moderately to severely impaired in patients with schizophrenia and their unaffected siblings and more than 80% of patients show significant impairment (Keefe and Fenton, 2007; Keefe and Harvey, 2012; Quee et al., 2014). This impairment group has an impact on outcomes such as occupational, social, clinical and economic functioning and emergent for treatment target. Research for pharmacological treatment is ongoing for improving cognitive function in schizophrenia but the results are not very convincing so far (Keefe and Harvey, 2012; Marder, 2006). The psycho-social intervention programs *e.g.* Cognitive Remediation (CR) is likely to be a safer way to improve cognitive functioning than pharmacologic treatment. The CR program produces modest improvements for patients with schizophrenia (Bora et al., 2009; Keefe and Harvey, 2012). Subtyping patients, especially severe cognitive impaired and severe symptom group, may be eligible for psychosocial interventions. A promising approach is cognitive adaption treatment (CAT), a method that takes into account the persistency of the cognitive impairments, while providing practical means to overcome these handicaps in real life situations (Quee et al., 2014).

Attention should be taken on subdomains of negative symptoms, especially when patients are severely ill. Evidence suggests there are deficits in anticipatory pleasure and defeatists beliefs from SA (Messinger et al., 2011; Foussias and Remington, 2008). Cognitive Behavioral Therapy (CBT) may be a suitable intervention for changing these beliefs. In addition, learning to anticipate on the experience of pleasurable events could possibly reduce SA. In this thesis, a significant association of ED with functioning was found after controlling for neuro-cognition. Therefore in chapter 4, we suggested that interventions targeting expressive skills such as Social Skills Training (Shean, 2009), could possibly improve ED. Though there is debate on the effectiveness of Social Skills Training, it has shown to be more effective in reducing negative symptoms than other psychosocial interventions (Turner et al., 2014).

Additionally, chapter 7 demonstrated that DUP was longer for patients being younger at onset of the psychotic disorder, for first generation immigrant patients and for male patients. Literature showed that longer DUP predicts worse symptoms at admission (Drake et al., 2000). Early implementation of intensive psychosocial intervention services may be necessary to shorten DUP, since the effect of DUP on recovery is greatest in the early stages of illness.

Future directions

Directions for statistical development

I elaborated the methodological aspects of several indices of hierarchical clustering technique using simulation studies in chapter 2. Overcoming some limitations of hierarchical and partitioning clustering, it would be of interest to use a multivariate finite mixture model, which classifies observations on the basis of probability estimated from Gaussian mixture modeling in cross-sectional studies. This method produces posterior distributions of all cluster parameters, proportions and cluster membership probabilities for all subjects (Kass and Raftery, 1995; Raftery and Dean, 2006). It would also be of interest to use clustering approach on multivariate non-normally distributed data (e.g. mixtures of other distributions).

Another focus would be extending clustering indices and apply those indices to trajectory modeling framework to confirm the number of subtypes (chapter 3 and 4). The performance of group-based trajectory analysis over growth-curve modeling and linear mixed effects modeling is less well known. Therefore, future studies may investigate the performance of group-based trajectory modeling on the selection of clusters using extensive simulations. So far, no one has investigated how well GBTM would detect the number of clusters and the trajectories in finite studies.

In longitudinal data, most methods deal with just one outcome variable over time to determine the number of clusters. It might be of a great interest to develop multivariate longitudinal clustering techniques, i.e. extending the group-based trajectory modeling to multivariate outcomes over time. For example, in this thesis, eight cognitive tests are formed into one composite measure to be able to identify the number of clusters and its trajectories. I would expect more subtle trajectories and different clusters if eight cognitive tests could be jointly incorporated into the trajectory modeling over time.

Another improvement area would be the integration of models for missingness with clustering or trajectory modeling. Throughout the thesis, I assumed the missing mechanisms to be MAR. It is not examined whether the missing mechanism is MAR or not, since it is impossible to verify if it would be MNAR. However, it would be highly interesting to model clustering technique in combination with MNAR models, but this requires more research on realistic missingness models in schizophrenia research.

Directions for clinical research

There is still abundant room for improving clinical research. Clustering cognitive functioning and symptoms longitudinally in such a way that all observations would be clustered at all time points separately, and transition between time points would be modeled by Markov transition matrices. This indicates that clustering of one time point will be affected by what chances at the adjacent time points (Franzen, 2008).

Phenotypically patients with schizophrenia and their unaffected siblings are cognitively heterogeneous. It may be beneficial to investigate the genetic effect on subtypes to confirm the true cognitive subtypes. This is an exciting, new chapter in research, now that polygenic risk scores have become available also for schizophrenia (Schizophrenia Working Group of the Psychiatric, Genomics Consortium, 2014). It would also be interesting to study whether the cognitive profiling approach is predicting functional and clinical outcomes over time.

It would be informative to use shorter follow-up times to study the effect of SA and ED on functioning, clinical outcomes and quality of life. In chapter 4, I demonstrated the effect of individual subtypes of SA and ED on outcomes over time. But symptoms of SA and ED tend to co-occur (Hartmann-Riemer et al., 2015) and may reinforce each other (Goekoop and Goekoop, 2014). Future studies should investigate in such interaction between subtypes of SA and ED, which may provide insight in the relationship with outcomes.

For future research it might be interesting to monitor psychotic symptoms of genetic high risk group more frequently (e.g. half-yearly) over time and study associations with neuro- and social cognition. Such analysis will be possible within the recently started Early Detection Study in the

Netherlands (see www.rgoc.nl/ontheroad). Additionally, one could study the discrete level of frequency and distress of psychotic experiences to observe the association with neuro- and social cognition instead of using continuum score of frequency and distress of psychotic experiences using generalized linear mixed effects modeling.

In chapter 6, I demonstrated that familial liability is one of the main determinants of multimorbidity which acts independently of the other major risk factors of age and gender. It was reported that the familial correlation between siblings and index patients were 11 to 16%. It should be essential to perform additional genetic analysis to distinguish between the effects of genetic liability and intra-familial environmental susceptibility on multimorbidity, applying polygenic risk scores for schizophrenia and other complex disorders in the same population.

Finally, it would be of interest to explore whether ethnicity plays part in the association between DUP and migration status. In addition, compelling literature indicates that there is an association between a shorter DUP and positive and negative symptoms, functional outcome and quality of life (Perkins et al., 2005; Boonstra et al., 2012). Therefore, a longitudinal study should be warranted to see whether an association between DUP and changes in functioning and symptoms exist.

Concluding remarks

The current thesis describes a number of new findings on the heterogeneity in cognitive functioning and clinical symptoms in schizophrenia patients and their unaffected siblings- with a special emphasis on the statistical approaches that were applied. In cross-sectional study-designs, hierarchical clustering with Duda and Hart (DH) index is the best approach to determine the number of clusters; however, model-based clustering approach would be preferable to confirm these clusters. In longitudinal studies, group-based trajectory modeling under finite mixture modeling is the best approach for summarizing and graphical representing of distinct trajectories.

Using these approaches, this thesis clearly underlined the validity of cognitive subtypes of patients and four subtypes of siblings, being stable and persistent over time and putting forward new clinical insights. Similarly, this thesis extent our knowledge on negative symptoms subtypes of SA and ED respectively, by demonstrating the performances of these subtypes being persistent on clinical and functional outcomes over time.

Another application of mixture of generalized linear mixed effects modeling on zero-inflated continuous outcome of psychotic experiences of siblings helped us to explain this outcome by the neuro-and social cognitive functioning, while at same time this yielded valuable insight in the clinical development of psychotic symptoms over a three-year time.

Psychotic experiences and symptoms may be on a continuum, reaching from the general non-ill population all the way to the schizophrenia spectrum disorders. Findings in patients and their non-affected siblings however confirm that for both cognition and negative symptoms heterogeneity exists and meaningful subtypes can be identified. These subtypes may provide new avenues to better understanding and more effectively treating people with psychotic disorders.

References

Agborsangaya, C.B., Lau, D., Lahtinen, M., Cooke, T., Johnson, J.A. (2012). Multimorbidity prevalence and patterns across socioeconomic determinants: a cross-sectional survey. *BMC public health* 12:201.

Allison, P.D. (2002). *Missing Data*. Thousand Oaks, CA: Sage Publications.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Washington, DC: American Psychiatric Publishing.

Appels, M.C., Sitskoorn, M.M., Westers, P., Lems, E., Kahn, R.S. (2003). Cognitive dysfunctions in parents of schizophrenic patients parallel the deficits found in patients. *Schizophrenia research* 63:285-293.

Bleuler, E. (1950). *Dementia Praecox or the Group of Schizophrenias*. New York: International Universities Press.

Boonstra, N., Klaassen, R., Sytema, S., Marshall, M., De Haan, L., Wunderink, L., Wiersma, D. (2012). Duration of untreated psychosis and negative symptoms--a systematic review and meta-analysis of individual patient data. *Schizophrenia research* 142:12-19.

Boonstra, N., Sterk, B., Wunderink, L., Sytema, S., De Haan, L., Wiersma, D. (2012). Association of treatment delay, migration and urbanicity in psychosis. *European psychiatry : the journal of the Association of European Psychiatrists* 27:500-505.

Bora, E., Veznedaroğlu, B., Vahip, S. (2016). Theory of mind and executive functions in schizophrenia and bipolar disorder: A cross-diagnostic latent class analysis for identification of neuropsychological subtypes. *Schizophrenia research* 176:500-505.

Bora, E., Yücel, M., Pantelis, C. (2009). Cognitive Impairment in Schizophrenia and Affective Psychoses: Implications for DSM-V Criteria and Beyond. *Schizophrenia bulletin* 36:36-42.

Carpenter, W.T., Kirkpatrick, B. (2015). Concepts and methods when considering negative symptom course. *Psychological medicine*1-2.

Commander, M.J., Cochrane, R., Sashidharan, S.P., Akilu, F., Wildsmith, E. (1999). Mental health care for Asian, black and white patients with non-affective psychoses: pathways to the psychiatric hospital, in-patient and after-care. *Social psychiatry and psychiatric epidemiology* 34:484-491.

Cotton, S.M., Wright, A., Harris, M.G., Jorm, A.F., McGorry, P.D. (2006). Influence of gender on mental health literacy in young Australians. *The Australian and New Zealand Journal of Psychiatry* 40:790-796.

Dawes, S.E., Jeste, D.V., Palmer, B.W. (2011). Cognitive profiles in persons with chronic schizophrenia. *Journal of clinical and experimental neuropsychology* 33:929-936.

De Hert, M., Correll, C.U., Bobes, J., Cetkovich-Bakmas, M., Cohen, D., Asai, I., Detraux, J., Gautam, S., Moller, H.J., Ndetei, D.M., Newcomer, J.W., Uwakwe, R., Leucht, S. (2011). Physical illness in patients with severe mental disorders. I. Prevalence, impact of medications and disparities in health care. *World psychiatry : official journal of the World Psychiatric Association (WPA)* 10:52-77.

Dollfus, S., Everitt, B., Ribeyre, J.M., Assouly-Besse, F., Sharp, C., Petit, M. (1996). Identifying subtypes of schizophrenia by cluster analyses. *Schizophrenia bulletin* 22:545-555.

Drake, R.J., Haley, C.J., Akhtar, S., Lewis, S.W. (2000). Causes and consequences of duration of untreated psychosis in schizophrenia. *The British journal of psychiatry : the journal of mental science* 177:511-515.

Ergul, C., UCok, A. (2015). Negative symptom subgroups have different effects on the clinical course of schizophrenia after the first episode: A 24-month follow up study. *European Psychiatry* 30:14-19.

Estivill-Castro, V. and Yang, J. (2000). Fast and Robust General Purpose Clustering Algorithms. In *PRICAI 2000 Topics in Artificial Intelligence: 6th Pacific Rim International Conference on Artificial Intelligence Melbourne, Australia, August 28 to September 1, 2000 Proceedings*, Mizoguchi, R. and Slaney, J. (eds), 208-218. Berlin, Heidelberg: Springer Berlin Heidelberg.

Fortin, M., Bravo, G., Hudon, C., Vanasse, A., Lapointe, L. (2005). Prevalence of multimorbidity among adults seen in family practice. *Annals of family medicine* 3:223-228.

Foussias, G., Siddiqui, I., Fervaha, G., Agid, O., Remington, G. (2014). Dissecting negative symptoms in schizophrenia: Opportunities for translation into new treatments. *Journal of psychopharmacology (Oxford, England)*1-11.

Foussias, G., Remington, G. (2008). Negative Symptoms in Schizophrenia: Avolition and Occam's Razor. *Schizophrenia bulletin* 36:359-369.

Fraley, C., Raftery, A.E. (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. *The Computer Journal* 41:578-588.

Fraley, C., Raftery, A.E. (2002). Model-Based Clustering, Discriminant Analysis, and Density Estimation. *Journal of the American Statistical Association* 97:611-631.

Franzen, J. (2008). Bayesian cluster analysis. *Department of Statistics, Stockholm University, Stockholm, Sweden*.

Fuchs, J., Busch, M., Lange, C., Scheidt-Nave, C. (2012). Prevalence and patterns of morbidity among adults in Germany. Results of the German telephone health interview survey German Health Update (GEDA) 2009. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz* 55:576-586.

Galderisi, S., Bucci, P., Mucci, A., Kirkpatrick, B., Pini, S., Rossi, A., Vita, A., Maj, M. (2013). Categorical and dimensional approaches to negative symptoms of schizophrenia: focus on long-term stability and functional outcome. *Schizophrenia research* 147:157-62.

Goekoop, R., Goekoop, J.G. (2014). A Network View on Psychiatric Disorders: Network Clusters of Symptoms as Elementary Syndromes of Psychopathology. *PLoS ONE* 9:e112734.

Goldstein, G., Allen, D.N., Seaton, B.E. (1998). A comparison of clustering solutions for cognitive heterogeneity in schizophrenia. *Journal of the International Neuropsychological Society : JINS* 4:353-362.

Goldstein, G. (1990). Neuropsychological heterogeneity in schizophrenia: A consideration of abstraction and problem-solving abilities. *Archives of Clinical Neuropsychology* 5:251-264.

Goldstein, G., Shemansky, W.J. (1995). Influences on cognitive heterogeneity in schizophrenia. *Schizophrenia research* 18:59-69.

Han, J., Kamber, M. (2001). *Data Mining: Concepts and Techniques*: Morgan Kaufmann Series in Data Management Systems.

Hartmann-Riemer, M., Hager, O.M., Kirschner, M., Bischof, M., Kluge, A., Seifritz, E., Kaiser, S. (2015). The association of neurocognitive impairment with diminished expression and apathy in schizophrenia. *Schizophrenia research* 169:427-32.

Hasler, G., Drevets, W.C., Gould, T.D., Gottesman, I.I., Manji, H.K. (2006). Toward Constructing an Endophenotype Strategy for Bipolar Disorders. *Biological psychiatry* 60:93-105.

Haviland, A.M., Jones, B.L., Nagin, D.S. (2011). Group-based Trajectory Modeling Extended to Account for Nonrandom Participant Attrition. *Sociological Methods & Research* 40:367-390.

Heinrichs, R.W., Awad, A.G. (1993). Neurocognitive subtypes of chronic schizophrenia. *Schizophrenia research* 9:49-58.

Heinrichs, R.W., Ruttan, L., Zakzanis, K.K., Case, D. (1997). Parsing Schizophrenia with Neurocognitive Tests: Evidence of Stability and Validity. *Brain and cognition* 35:207-224.

Jablensky, A. (2006). Subtyping schizophrenia: implications for genetic research. *Molecular psychiatry* 11:815-836.

Jacobi, F., Wittchen, H.U., Holting, C., Hofler, M., Pfister, H., Muller, N., Lieb, R. (2004). Prevalence, co-morbidity and correlates of mental disorders in the general population: results from the German Health Interview and Examination Survey (GHS). *Psychological medicine* 34:597-611.

Jones, B.L., Nagin, D.S., Roeder, K. (2001). A SAS Procedure Based on Mixture Models for Estimating Developmental Trajectories. *Sociological Methods & Research* 29:374-393.

Joyce, E.M., Hutton, S.B., Mutsatsa, S.H., Barnes, T.R. (2005). Cognitive heterogeneity in first-episode schizophrenia. *The British journal of psychiatry : the journal of mental science* 187:516-522.

Joyce, E.M., Roiser, J.P. (2007). Cognitive heterogeneity in schizophrenia. *Current opinion in psychiatry* 20:268-272.

Jung, T., Wickrama, K.A.S. (2008). An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling. *Social and Personality Psychology Compass* 2:302-317.

Kass, R.E., Raftery, A.E. (1995). Bayes Factors. *Journal of the American Statistical Association* 90:773-795.

Keefe, R.S., Fenton, W.S. (2007). How should DSM-V criteria for schizophrenia include cognitive impairment?. *Schizophrenia bulletin* 33:912-920.

Keefe, R.S., Harvey, P.D. (2012). Cognitive impairment in schizophrenia. *Handbook of Experimental Pharmacology* (213):11-37. doi:11-37.

Korver, N., Quee, P.J., Boos, H.B.M., Simons, C.J.P., de Haan, L., GPOUPinvestigators, Investigators, G. (2012). Genetic Risk and Outcome of Psychosis (GROUP), a multi-site longitudinal cohort study focused on gene-environment interaction: objectives, sample characteristics, recruitment and assessment methods. *International Journal of Methods in Psychiatric Research* 21:205-221.

Krabbendam, L., Marcelis, M., Delespaul, P., Jolles, J., van Os, J. (2001). Single or multiple familial cognitive risk factors in schizophrenia?. *American Journal of Medical Genetics* 105:183-188.

Kremen, W.S., Seidman, L.J., Pepple, J.R., Lyons, M.J., Tsuang, M.T., Faraone, S.V. (1994). Neuropsychological risk indicators for schizophrenia: a review of family studies. *Schizophrenia bulletin* 20:103-119.

Liemburg, E., Castelein, S., Stewart, R., van, d.G., Aleman, A., Knegtering, H. (2013). Two subdomains of negative symptoms in psychotic disorders: established and confirmed in two large cohorts. *Journal of psychiatric research* 47:718-25.

Liemburg, E., Castelein, S., Stewart, R., van, d.G., Aleman, A., Knegtering, H. (2013). Two subdomains of negative symptoms in psychotic disorders: established and confirmed in two large cohorts. *Journal of psychiatric research* 47:718-25.

Little, R.J.A. (1995). Modeling the Drop-Out Mechanism in Repeated-Measures Studies. *Journal of the American Statistical Association* 90:1112-1121.

Little, R.J.A. (1993). Pattern-Mixture Models for Multivariate Incomplete Data. *Journal of the American Statistical Association* 88:125-134.

Lykouras, L., Oulis, P., Daskalopoulou, E., Psarros, K., Christodoulou, G.N. (2001). Clinical subtypes of schizophrenic disorders: a cluster analytic study. *Psychopathology* 34:23-28.

Marder, S.R. (2006). Initiatives to promote the discovery of drugs to improve cognitive function in severe mental illness. *The Journal of clinical psychiatry* 67:e03.

Markova, I.S., Berrios, G.E. (1995). Mental symptoms: are they similar phenomena? The problem of symptom heterogeneity. *Psychopathology* 28:147-157.

Marshall, M., Lewis, S., Lockwood, A., Drake, R., Jones, P., Croudace, T. (2005). Association between duration of untreated psychosis and outcome in cohorts of first-episode patients: a systematic review. *Archives of General Psychiatry* 62:975-983.

Meijer, J., Simons, C.J., Quee, P.J., Verweij, K., GROUP Investigators. (2012). Cognitive alterations in patients with non-affective psychotic disorder and their unaffected siblings and parents. *Acta Psychiatrica Scandinavica* 125:66-76.

Messinger, J.W., Tremeau, F., Antonius, D., Mendelsohn, E., Prudent, V., Stanford, A.D., Malaspina, D. (2011). Avolition and expressive deficits capture negative symptom phenomenology: implications for DSM-5 and schizophrenia research. *Clinical psychology review* 31:161-8.

Michiels, B., Molenberghs, G., Bijnens, L., Vangeneugden, T., Thijs, H. (2002). Selection models and pattern-mixture models to analyse longitudinal quality of life data subject to drop-out. *Statistics in medicine* 21:1023-1041.

Miller, G.A., Rockstroh, B. (2013). Endophenotypes in Psychopathology Research: Where Do We Stand?. *Annual Review of Clinical Psychology* 9:177-213.

Milligan, G.W., Cooper, M.C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika* 50:159-179.

Muthen, B., Muthen, L.K. (2000). Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes. *Alcoholism, Clinical and Experimental Research* 24:882-891.

Muthen, B., Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics* 55:463-469.

Nagin, D.S. (2014). Group-based trajectory modeling: an overview. *Annals of Nutrition & Metabolism* 65:205-210.

Nerhus, M., Berg, A.O., Haram, M., Kvitland, L.R., Andreassen, O.A., Melle, I. (2015). Migrant background and ethnic minority status as predictors for duration of untreated psychosis. *Early intervention in psychiatry* 9:61-65.

Norman, R., Manchanda, R., Harricharan, R., Northcott, S. (2015). The course of negative symptoms over the first five years of treatment: Data from an early intervention program for psychosis. *Schizophrenia research*in press-in press.

Nuyen, J., Schellevis, F.G., Satariano, W.A., Spreeuwenberg, P.M., Birkner, M.D., van den Bos, G.A., Groenewegen, P.P. (2006). Comorbidity was associated with neurologic and psychiatric diseases: a general practice-based controlled study. *Journal of clinical epidemiology* 59:1274-1284.

Oreski, I., Jakovljevic, M., Aukst-Margetic, B., Orlic, Z.C., Vuksan-Cusa, B. (2012). Comorbidity and multimorbidity in patients with schizophrenia and bipolar disorder: similarities and differencies. *Psychiatria Danubina* 24:80-85.

Perkins, D.O., Gu, H., Boteva, K., Lieberman, J.A. (2005). Relationship between duration of untreated psychosis and outcome in first-episode schizophrenia: a critical review and meta-analysis. *The American Journal of Psychiatry* 162:1785-1804.

Quee, P.J., Alizadeh, B.Z., Aleman, A., van den Heuvel, E.R., GROUP Investigators. (2014). Cognitive subtypes in non-affected siblings of schizophrenia patients: characteristics and profile congruency with affected family members. *Psychological medicine* 44:395-405.

Quee, P.J., Stiekema, A.P., Wigman, J.T., Schneider, H., van der Meer, L., Maples, N.J., van den Heuvel, E.R., Velligan, D.I., Bruggeman, R. (2014). Improving functional outcomes for schizophrenia patients in the Netherlands using Cognitive Adaptation Training as a nursing intervention - A pilot study. *Schizophrenia research* 158:120-125.

Raftery, A.E., Dean, N. (2006). Variable Selection for Model-Based Clustering. *Journal of the American Statistical Association* 101:168-178.

Ritsner, M. S. and Gottesman, I. I. (2011). The Schizophrenia Construct After 100 Years of Challenges. In *Handbook of Schizophrenia Spectrum Disorders, Volume I: Conceptual Issues and Neurobiological Advances*, Ritsner, M. S. (ed), 1-44. Dordrecht: Springer Netherlands.

Rokach, L. and Oded, M. (2005). Clustering methods. In *Data mining and knowledge discovery handbook*321-352. US: Springer.

Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys*. New York: J. Wiley & Sons.

Sautter, F.J., McDermott, B.E., Cornwell, J., Johnson, J., Borges, A., Wilson, A.F., Vasterling, J.J., Foundas, A.L. (1995). A preliminary study of the neuropsychological heterogeneity of familial schizophrenia. *Schizophrenia research* 18:1-7.

Schafer, J.L. (1997). *Analysis of Incomplete Multivariate Data*. New York: Chapman and Hall.

Schimmelmann, B.G., Huber, C.G., Lambert, M., Cotton, S., McGorry, P.D., Conus, P. (2008). Impact of duration of untreated psychosis on pre-treatment, baseline, and outcome characteristics in an epidemiological first-episode psychosis cohort. *Journal of psychiatric research* 42:982-990.

Schizophrenia Working Group of the Psychiatric,Genomics Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511:421-427.

Seaton, B.E., Allen, D.N., Goldstein, G., Kelley, M.E., van Kammen, D.P. (1999). Relations between cognitive and symptom profile heterogeneity in schizophrenia. *The Journal of nervous and mental disease* 187:414-419.

Seaton, B.E., Goldstein, G., Allen, D.N. (2001). Sources of Heterogeneity in Schizophrenia: The Role of Neuropsychological Functioning. *Neuropsychology review* 11:45-67.

Shah, N.H., Hipwell, A.E., Stepp, S.D., Chang, C.H. (2014). Measures of discrimination for latent group-based trajectory models. *Journal of applied statistics* 42:1-11.

Shean, G.D. (2009). Evidence-based psychosocial practices and recovery from schizophrenia. *Psychiatry* 72:307-320.

Snitz, B.E., Macdonald, A.W.,3rd, Carter, C.S. (2006). Cognitive deficits in unaffected first-degree relatives of schizophrenia patients: a meta-analytic review of putative endophenotypes. *Schizophrenia bulletin* 32:179-194.

Sterk, B., Slief, E.M., Blankers, M., Linszen, D.H., de Haan, L. (2010). Duration of untreated psychosis and ethnicity. *Schizophrenia research* 124:238-239.

Stiekema, A.P.M., Liemburg, E.J., van, d.M., Castelein, S., Stewart, R., van Weeghel, J., Aleman, A., Bruggeman, R. (2016). Confirmatory factor analysis and differential relationships of the two subdomains of negative symptoms in chronically ill psychotic patients. *PLoS ONE* 11:e0149785-e0149785.

Thijs, H., Molenberghs, G., Michiels, B., Verbeke, G., Curran, D. (2002). Strategies to fit pattern-mixture models. *Biostatistics* 3:245-265.

Tibshirani, R., Walther, G., Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society.Series B, Statistical methodology* 63:411-423.

Tooze, J.A., Grunwald, G.K., Jones, R.H. (2002). Analysis of repeated measures data with clumping at zero. *Statistical methods in medical research* 11:341-355.

Turner, D.T., van, d.G., Karyotaki, E., Cuijpers, P. (2014). Psychological Interventions for Psychosis :. *American Journal of Psychiatry* 171:523-538.

van Buuren, S. (2007). Multiple Imputation of Discrete and Continuous Data by Fully Conditional Specification. *Statistical methods in medical research* 16:219-242.

Verbeke, G., Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer series in statistics.

Willadsen, T.G., Bebe, A., Koster-Rasmussen, R., Jarbol, D.E., Guassora, A.D., Waldorff, F.B., Reventlow, S., Olivarius Nde, F. (2016). The role of diseases, risk factors and symptoms in the definition of multimorbidity - a systematic review. *Scandinavian journal of primary health care* 34:112-121.

Wolff, G., Pathare, S., Craig, T., Leff, J. (1996). Community knowledge of mental illness and reaction to mentally ill people. *The British journal of psychiatry: the journal of mental science* 168:191-198.

Addendums

Nederlandse Samenvatting

Mensen met psychotische stoornissen laten grote verschillen zien op het gebied van symptomen (zoals positieve en negatieve symptomen en cognitieve beperkingen) en de presentatie daarvan. Deze heterogeniteit wordt al lang onderkend, maar desondanks vaak genegeerd, zowel in de klinische praktijk als in het onderzoek. Het hoofddoel van dit proefschrift is dan ook het ontrafelen van de heterogeniteit in cognitief functioneren en van de klinische symptomen bij mensen met een psychotische stoornis en hun niet-zieke broers en zusters, gebruikmakend van cross-sectionele en longitudinale gegevens. De nadruk ligt bij dit alles vooral op de statistische benaderingen die zijn gebruikt.

Eén manier om heterogeniteit te onderzoeken bij patiënten met eenzelfde diagnose is door datareductie technieken, zoals klassieke clusteranalyse, en ”group-based trajectory models (GBTM)” toe te passen, om hiermee homogene subtypes te vormen. De studies in dit proefschrift zijn allemaal uitgevoerd binnen het GROUP (Genetic Risk and Outcome of Psychosis) project. Dit is een grote longitudinale multicenter cohort studie in Nederland en België met een uitgebreide eerste meting, gevolgd door nog twee metingen na drie en na zes jaar.

Hoofdstuk 1 geeft een inleiding in de schizofrenie spectrumstoornissen, de klinische heterogeniteit, en de statistische technieken die geschikt zijn om heterogeniteit te kunnen bestuderen. Er is aanzienlijk bewijs voor significante heterogeniteit in cognitief functioneren en klinische symptomen.

In **Hoofdstuk 2** wordt een overzicht gegeven van de meest belovende indices (methodes) om het juiste aantal clusters te berekenen op basis van een hiërarchische clustertechniek, gebruikmakend van de Ward’s agglomeratiemethode. Deze indices zijn toegepast op een achttal cognitieve uitkomstvariabelen bij patiënten met schizofrenie (“case study”) en binnen simulatiedata. Van de 14 onderzochte indices bleken Duda en Hart (DH), Hartigan (H) en Gap/pc de best voorspellende indices. De DH index was het meest consistent, terwijl Gap/pc en WGap/pc in staat bleken te zijn om de aanwezigheid van meerdere clusters aan te tonen.

Hoofdstuk 3 toont cognitieve subtypes van patienten met schizofrenie en hun broers/zussen aan, op basis van een gemiddelde cognitiescore, gebruikmakend van GBTM. Er werden bij patienten vijf subtypes gevonden en bij hun broers en zussen vier subtypes, die stabiel bleken te zijn over de periode van zes jaar. Met het oog op de familiestructuur van de data (patiënten en hun broers en zussen behoren tot dezelfde familie) werd een “clustered multinominal logistic regression” uitgevoerd, met de vraag of subtypes van patiënten de subtypes van hun broers en zussen zouden voorspellen. Dit bleek significant te voorspellen te zijn. Hoe slechter het cognitieve profiel van de patiënt, hoe beter dit voorspelde wat het subtype van de broers en zussen was (OR 10.07, 95% CI 4.15–24.44). Dit bleek ook voor de groep met gematigd cognitief functioneren te gelden (OR 5.7, 95% CI 2.77–11.70). De intraclass correlatie tussen de index-patiënten en hun broers en zussen verklaarde 27% van de totale variatie.

Hoofdstuk 4 beschrijft de toepassing van GBTM bij het bepalen van homogene groepen van patiënten, gebaseerd op het beloop van de negatieve subdomeinen van symptomen, zoals *social amotivation* (SA) en *expressive deficit* (ED). Ook werd hier de vraag gesteld of deze homogene

groepen bijdragen aan meer inzicht in de subdomeinen van negatieve symptomen, en of zij van belang zijn voor het functioneren en kwaliteit van leven. Voor het beantwoorden van deze tweede vraag werd gebruik gemaakt van (generalized) linear mixed models. Er bleek een significante heterogeniteit te bestaan in het beloop van de negatieve symtpomen, suggererend dat negatieve symptomen minder stabiel zijn dan altijd verondersteld werd. Subgroepen binnen de SA- en ED-groep lieten een verschillend beloop zien. Het klinisch belang van de subtypes werd onderstreept door hun verschillende relaties met het beloop en de ernst van de uitkomstmaten (functioneren en kwaliteit van leven).

In **Hoofdstuk 5** gaat het om de toepassing van een " mixture of generalized linear mixed effects models", die gebruikt werd om de ontwikkeling van psychotische belevingen van broers en zussen van patiënten te beschrijven en de relatie met neuro- en sociale cognitie. Slechter verbaal leren voorspelde het optreden van psychotische belevingen en de daarmee gepaard gaande stress drie jaar later. Bovendien bleken Theory of Mind taken (hinting task) geassocieerd te zijn met een afname van psychotische belevingen na drie jaar. In de Hoofdstukken 4 en 5 werd ook de techniek van Multiple Imputation toegepast.

De heterogeniteit van somatische aandoeningen en klachten van patiënten, hun broers en zussen, en gezonde controle personen wordt beschreven in **Hoofdstuk 6**. Het effect van de familiaire gevoeligheid voor psychosen, naast het effect van geslacht en leeftijd, werd onderzocht met generalized linear mixed effects modelling. Familiaire kwetsbaarheid bleek een significante predictor van multimorbiditeit te zijn.

Hoofdstuk 7 bestudeert de risicofactoren voor de duur van onbehandelde psychose (DOP) met behulp van ordinale logistische regressie, samen met het "Cox-proportional hazard" regressiemodel. DOP geeft de tijd weer tussen het ontstaan van de psychotische klachten en het begin van de behandeling. Uit beide analyses bleken migratiestatus, leeftijd van ontstaan van de klachten en geslacht significant samen te hangen met de DOP. Mannelijke, eerste generatie migranten met een jonge leeftijd van ontstaan van de klachten, bleken het hoogste risico te hebben op een lange DOP.

Tenslotte worden in **Hoofdstuk 8** de belangrijkste bevindingen en de gebruikte statistische benaderingen samengevat en bediscussieerd. Ook wordt het wetenschappelijk belang van de bevindingen besproken, samen met de nieuwe methodologische overwegingen en aanbevelingen voor toekomstig onderzoek.

Samenvattend, voor cross-sectionele data blijkt hiërarchisch clustering volgens de Duda en Hart index de beste benadering te zijn om het aantal clusters te bepalen. Echter, een model-based clustering benadering zou de voorkeur hebben om deze clusters te bevestigen. In longitudinale studies GBTM met finite-mixture modelling voldoet het beste als benadering om omschreven trajectories in beeld te brengen.

Gebruikmakend van deze benaderingswijzen, laat dit proefschrift duidelijk de validiteit van de cognitieve subtypes zien voor patiënten en broers en zussen, die stabiel zijn over de tijd. Daarmee worden ook nieuwe klinische inzichten verschaft. Ook draagt dit proefschrift bij aan onze kennis van negatieve symptomen en de subdomeinen SA en ED, door te laten zien dat deze subtypes persistent zijn en samen hangen met klinische uitkomsten.

De toepassing van een mixture of generalized linear mixed effects modelling bij een “zero-inflated” continue uitkomstmaat van psychotische belevingen bij broers en zussen van patiënten, liet het effect zien van deze parameters op neuro- en sociaal cognitief functioneren. Tegelijkertijd leverde dit belangrijke inzichten op over de klinische ontwikkeling van psychotische symptomen over drie jaar.

Psychotische belevingen en symptomen vormen wellicht een continuüm reikend van de algemene bevolking tot aan mensen met een schizofrenie spectrumstoornis. De bevindingen van dit proefschrift laten echter zien dat voor patienten en hun broers en zussen heterogeniteit bestaat op het gebied van cognitie en negatieve symptomen, en dat betekenisvolle subtypes gevonden kunnen worden. Deze subtypes kunnen tot nieuwe wegen leiden, met meer inzicht en een betere behandeling van mensen met een psychotische stoornis.

List of publications

- **Islam, M.A.,** Khan, M.F.H., Quee, P.J., Snieder, H., van den Heuvel, E.R., Alizadeh, B.Z. and GROUP Investigators (2017). Familial liability to psychosis is a risk factor for multimorbidity in People with psychotic disorders and their unaffected siblings. In press*, European Psychiatry*.
- van Os, J., van der Steen, Y., **Islam, M.A.,** Gülöksüz, S., Rutten, B.P., Simons, C.J., and GROUP Investigators (2017). Evidence that polygenic risk for psychotic disorder is expressed in the domain of neurodevelopment, emotion regulation and attribution of salience, *Psychological Medicine*, 1-17, DOI:10.1017/S0033291717000915.
- Stiekema, A.P.M., **Islam, M.A**., Lieburg, E.J., Castelein, S., van den Heuvel, E.R., van Weeghel, J., Aleman, A., Bruggeman, R., van der Meer, L., and GROUP Investigators (2017). Long-term course of negative symptom subdomains and relationship with outcome in patients with a psychotic disorder. In revision, *Schizophrenia Research.*
- **Islam, M.A**[*]., Pijnenborg, G.H.M[*]., van den Heuvel, E.R., van der Gaag, Mark., Bruggeman, R, Bartels-Velthuis, A.A., de Vries, S., and GROUP Investigators (2017). The predictive value of neuro and social cognition for the development of psychotic experiences in sibling of people with psychotic disorders ([*]Joint 1st author). *Submitted to Journal of Neuropsychology.*
- **Islam, M.A**., van Es, F.D., Habtewold, T.D, Quee, P.J., van den Heuvel, E.R., Alizadeh, B.Z., Bruggeman, R., and GROUP Investigators. (2017). Long-term cognitive trajectories in patients with schizophrenia and their unaffected siblings. *Submitted to Psychological Medicine*.
- Habtewold, T.D., **Islam, M.A**., Radi, Y.T., and Tegegne, B.S. (2016). Comorbidity of depression and diabetes: an application of biopsychosocial model, *International Journal of Mental Health Systems*, 10:74. Doi: 10.1186/s13033-016-0106-2
- Lieburg, E.J., Bruins, J., van Beveren, N., **Islam, M.A**., Alizadeh, B.Z., and GROUP investigators (2016). Cannabis and a lower BMI in psychosis: What is the role of AKT1? *Schizophrenia Research*, **176**, 95-99.
- **Islam, M.A**., Hossain, M.T., Sarwar, G., Kawsar, L.A., Smrity, L.A., Alam, M.A., Hossain, K.H., and Bhuia, M.R. (2016). Is Socio-economic Condition of Manipuri Tribe Improving? - A two Phase Comparison Using Income Distribution Model, *International Journal of Statistics and Applications*, **6(3)**, 138-144. DOI: 10.5923/j.statistics.20160603.06
- **Islam, M.A**., Hossain, M.T., Sarwar, G., Kawsar, L.A., Smrity, L.A., Alam, M.A., Hossain, K.H., and Bhuia, M.R. (2016). Structural Equation Modeling to Assess the Impact of Socio-Demographic Variables on Fertility of Ethnic *Manipuri* Women, *Biometrics & Biostatistics International Journal*, **3(5)**, 00078. DOI: 10.15406/bbij.2016.03.00078
- Ferdous, M., Friedrich, A., Grundmann, H., de Boer, R., Croughs, P., **Islam, M.A**., Kluytmans-van den Bergh., Marjolein, K.S.A., and Rossen, J. (2016). Molecular characterization and phylogeny of Shiga toxin-producing E. coli (STEC) isolates obtained from two Dutch regions using Whole Genome Sequencing, *Clinical Microbiology and Infection*, **22(7)**, 642.e1–642.e9
- **Islam, M.A**., Alizadeh, B.Z., van den Heuvel, E.R., and GROUP Investigators (2015). A comparison of indices for identifying the number of clusters in hierarchical clustering: A study on cognition in schizophrenia patients, *Communication in Statistics-Case studies, data analysis and applications*, **1(2)**, 98-113.

- Apeldoorn, S.Y., Sterk, B., van den Heuvel, E.R., Schoevers, R.A., **Islam, M.A**., and GROUP Investigators (2014). Factors contributing to the duration of untreated psychosis, Schizophrenia Research, **158(1-3)**, 76-81. http://dx.doi.org/10.1016/j.schres.2014.07.002
- Quee, P.J., Meijer, J.H., **Islam, M.A**., Aleman, A., Alizadeh, B.Z., Meijer, C.J., van den Heuvel, E.R. and GROUP Investigators (2014). Premorbid Adjustment Profiles in Psychosis and the Role of Familial Factors, *Journal of Abnormal Psychology,* **123(3)**, 578-87.
- Kabir, A., Ali, R., M. Shahidul Islam, M.S., Kawsar, L.A., and **Islam, M.A.** (2009). A Comparison of the Regional Variations of Fertility in Bangladesh, *International Quarterly of Community Health Education, New York,* **29(3),** 275-291.
- Kawsar, L.A., Kabir, A., and **Islam, M.A.** (2009). Sex Preference and Contraceptive Norms among Different Regions in Bangladesh, *International Quarterly of Community Health Education*, *New York*, **29(2),** 183-197.
- Bhuia, M.R., Hossain, M.K., Hossain, M.F., **Islam**, **M.A**., Tofayel. A., and Jobayer, S.M. (2009). Knowledge, Attitude and Community views of Forced Marriage-A Study in Sylhet division. *Jahangirnagar University Journal of Science*, **32(1),** 103-113.
- Das, Sumonkanti., Hossain, M.Z., and **Islam, M.A.** (2008). Predictors of Child Chronic Malnutrition in Bangladesh, *Proc. Pakistan Academy of Science,* **45(3),** 137-155.

Co-authors

André Aleman: Department of Neuroscience, University Medical Center Groningen, University of Groningen, The Netherlands

Annemarie P. M. Stiekema: School for Mental Health and Neuroscience, Maastricht University Medical Centre, Maastricht, The Netherlands

Behrooz Z. Alizadeh: Department of Epidemiology, University Medical Center Groningen, University of Groningen, The Netherlands

Bouke Sterk: Department of Psychiatry, Academic Medical Centre, University of Amsterdam, the Netherlands

Edith J. Liemburg: University Center for Psychiatry, Rob Giel Research Center, University of Groningen, University Medical Center Groningen, The Netherlands

Edwin R. van den Heuvel: Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands

Frank D. van Es: University Center for Psychiatry, Rob Giel Research Center, University of Groningen, University Medical Center Groningen, The Netherlands

Gerdina H.M. Pijnenborg: Department of Clinical Psychology, University of Groningen, The Netherlands

Harold Snieder: Department of Epidemiology, University Medical Center Groningen, University of Groningen, The Netherlands

Jaap van Weeghel: Parnassia Group, Dijk en Duin Mental Health Center, Castricum, Department of TRANZO, Tilburg School of Social and Behavioral Sciences, Tilburg University, The Netherlands

Lisette van der Meer: University Center for Psychiatry, Rob Giel Research Center, University of Groningen, University Medical Center Groningen, The Netherlands

Mark van der Gaag: Department of Clinical Psychology, VU University and EMGO Institute of health and Care Research, Amsterdam, The Netherlands

M. Feroz Hayat Khan: Consultant, WHO, Bangladesh, and University Center for Psychiatry, Rob Giel Research Center, University of Groningen, University Medical Center Groningen, The Netherlands

Piotr J. Quee: University Psychiatric Centre (UPC), KU Leuven, campus Kortenberg, Leuven, Belgium

Richard Bruggeman: University Center for Psychiatry, Rob Giel Research Center, University of Groningen, University Medical Center Groningen, The Netherlands

Robert A. Schoevers: University Center for Psychiatry, University of Groningen, University Medical Center Groningen, The Netherlands

Sanne Y. Apeldoorn: University Center for Psychiatry, University of Groningen, University Medical Center Groningen, The Netherlands

Stefanie M. de Vries: Department of Clinical and Developmental Neuropsychology, University of Groningen, The Netherlands

Stynke Castelein: University Center for Psychiatry, Rob Giel Research Center, University of Groningen, University Medical Center Groningen, Lentis Psychiatric Institute, The Netherlands

S. Gülöksüz: Department of Psychiatry and Psychology, Maastricht University Medical Center, School for Mental Health and Neuroscience, Maastricht, The Netherlands

Tesfa D. Habtewold: Department of Epidemiology, University Medical Center Groningen, University of Groningen, The Netherlands

Y. van der Steen: Department of Psychiatry and Psychology, Maastricht University Medical Center, School for Mental Health and Neuroscience, Maastricht, The Netherlands

Genetic Risk and Outcome of Psychosis (GROUP) investigators: Behrooz Z. Alizadeh[a], Agna A. Bartels-Velthuis[a], Nico J. van Beveren[b,c,d], Richard Bruggeman[a], Wiepke Cahn[e], Lieuwe de Haan[f], Philippe Delespaul[g], Carin J. Meijer[f], Inez Myin-Germeys[h], Rene S. Kahn[e], Frederike Schirmbeck[f], Claudia J.P. Simons[g,j], Neeltje E. van Haren[e], Jim van Os[g,j], Ruud van Winkel[g,h]

[a]University of Groningen, University Medical Center Groningen, University Center for Psychiatry, Groningen, The Netherlands
[b]Antes Center for Mental Health Care, Rotterdam, The Netherlands
[c]Erasmus MC, Department of Psychiatry, Rotterdam, The Netherlands
[d]Erasmus MC, Department of Neuroscience, Rotterdam, The Netherlands
[e]University Medical Center Utrecht, Department of Psychiatry, Brain Centre Rudolf Magnus, Utrecht, The Netherlands
[f]Academic Medical Center, University of Amsterdam, Department of Psychiatry, Amsterdam, The Netherlands
[g]Maastricht University Medical Center, Department of Psychiatry and Psychology, School for Mental Health and Neuroscience, Maastricht, The Netherlands
[h]KU Leuven, Department of Neuroscience, Research Group Psychiatry, Center for Contextual Psychiatry, Leuven, Belgium
[i]GGzE, Institute for Mental Health Care Eindhoven and De Kempen, Eindhoven, The Netherlands
[j]King's College London, King's Health Partners, Department of Psychosis Studies, Institute of Psychiatry, London, United Kingdom

List of RGOc dissertations

Cognitive self-therapy. A contribution to long-term treatment of depression and anxiety.
Peter C.A.M. den Boer
RGOc-reeks, nummer 12, Groningen 2006, 198 pagina's

Antipsychotic Treatment Strategies and Early Course of First Episode Psychosis.
Lex Wunderink
RGOc-reeks, nummer 13, Groningen 2006, 184 pagina's

Economic evaluation in mental healthcare: Assessing the cost-effectiveness of interventions for patients with major depressive disorder or schizophrenia in the context of the Dutch healthcare system
Dennis Stant
RGOc-reeks, nummer 15, Groningen 2007, 144 pagina's

Bereaved by Suicide: Family-based cognitive-behavior therapy to prevent adverse health consequences among first-degree relatives and spouses
Marieke de Groot
RGOc-reeks, nummer 17, Groningen 2008, 215 pagina's

Guided peer support groups for psychosis. A randomized controlled trial.
Stynke Castelein
RGOc-reeks, nummer 18, Groningen, 2009, 147 pagina's.

Pharmacogenetics of Antipsychotic-Induced Parkinsonism and Tardive Dyskinesia: a Focus on African-Caribbeans and Slavonic Caucasians.
Asmar Al Hadithy
RGOc-reeks nummer 21, Groningen, 2009, 171 pagina's.

Drug safety in patients with psychotic disorders.
Susanne Gabriele Schorr
RGOc-reeks nummer 23, Groningen, 2010, 125 pagina's.

Desired and undesired effects of antipsychotic treatment from a patients perspective: the psychometric evaluation of a self-rating instrument.
Hugo Arnout Wolters
RGOc-reeks nummer 24, Groningen, 2010, 137 pagina's.

Auditory Hallucinations in Childhood.
Agna Bartels
RGOc-reeks nummer 26, Groningen, 2011, 152 pagina's.

Personalized pharmacotherapy of psychosis: Clinical and pharmacogenetic approaches.
Jelle Vehof
RGOc-reeks nummer 27, Groningen, 2011, 181 pagina's.

Early detection of psychosis; why should we care?
Nynke Boonstra
RGOc-reeks nummer 29, Groningen, 2011, 147 pagina's.

Recovery in first episode psychosis: role of neurocognition, antipsychotics and cannabis.
Gunnar Faber
RGOc-reeks nummer 30, Groningen, 2011, 152 pagina's.

Successful community living: a 'Utopia'? A survey of people with severe mental illness in Dutch Regional Institutes for Residential Care.
Charlotte de Heer
RGOc-reeks nummer 31, Groningen, 2012, 145 pagina's.

Cognitive functioning in schizophrenia: structure and clinical correlates.
Piotr J. Quee
RGOc-reeks nummer 32, Groningen, 2013, 165 pagina's.

Patients in the driver's seat: a role for e-mental health?
Lian van der Krieke
RGOc-reeks nummer 33, Groningen, 2014, 236 pagina's.

START-ing risk assessment and shared care planning in out-patient forensic psychiatry. Results from a cluster randomized controlled trial.
Nadine Troquete
RGOc-reeks nummer 35, Groningen, 2014, 231 pagina's.

Metabolic risk in people with psychotic disorders: No mental health without physical health
Jojanneke Bruins
RGOc-reeks nummer 36, Groningen, 2016, 193 pagina's.

List of SHARE dissertations

Research Institute SHARE

This thesis is published within the **Research Institute SHARE** (Science in Healthy Ageing and healthcaRE) of the University Medical Center Groningen / University of Groningen.

Further information regarding the institute and its research can be obtained from our internet site: http://www.share.umcg.nl/

More recent theses can be found in the list below.
((co-) supervisors are between brackets)

2017

Monden R
Deconstruction depression; a 3d perspective
(prof P de Jonge, dr KJ Wardenaar, dr A Stegeman)

Daud NAA
Paving ways for personalizing drug therapy during pregnancy; a focus on the risk of drug teratogenicity
(prof B Wilffert, dr JEH Bergman)

Spoorenberg SLW
Embracing the perspectives of older adults in organising and evaluating person-centred and integrated care
(prof SA Reijneveld, prof HPH Kremer, dr K Wynia)

Uittenbroek RJ
Impact of person-centered and integrated care for community-living older adults on quality of care and service use and costs
(prof SA Reijneveld, prof HPH Kremer, dr K Wynia)

Folbert E
Geriatric traumatology; the effectiveness of integrated orthogeriartric treatment on 1-year outcome in frail elderly with hip fracture
(prof JPJ Slaets, prof HJ ten Duis, dr JH Hegeman)

Panman CMCR & Wiegersma M
Pelvic organ prolapse; conservative treatments in primary care
(prof MY Berger, dr JH Dekker)

Postema SG
Upper limb absence; effects on body functions and structures, musculoskeletal complaints and functional capacity
(prof CK van der Sluis, prof MF Reneman, dr RM Bongers)

Adrichem EJ van
Physical activity in recipients of solid organ transplantation
(prof CP van der Schans, prof PU Dijkstra, dr R Dekker)

Luten KA
Development and evaluation of a community-based approach to promote healh-related behavior among older adults in a socioeconomically disadvantaged community
(prof A Dijkstra, prof SA Reijneveld, dr AF de Winter)

Setiawan D
HPV vaccination in Indonesia; a health-economic & comparative perspective
(prof MJ Postma, prof B Wilffert, dr JA Thobari)

Sluis A van der
Risk factors for injury in talented soccer and tennis players; a maturation-driven approach
(prof C Visscher, dr MT Elferink-Gemser, dr MS Brink)

Bouwmans MEJ
A sad day's night; the dynamic role of sleep in the context of major depression.
(prof P de Jonge, prof AJ Oldehinkel)

Bakker M
Challenges in prenatal screening and diagnosis in the Netherlands
(prof CM Bilardo, dr E Birnie)

Annema-de Jong JH
What's on your mind? Emotions and perceptions of liver transplant candidates and recipients
(prof AV Ranchor, prof PF Roodbol, prof RJ Porte)

2016

Does HTJ van der
Enhancing performance & preventing injuries in team sport players
(prof KAPM Lemmink, prof C Visscher, dr MS Brink)

Veldman K
Mental health from a life course perspective; the transition from school to work
(prof U Bültmann, prof SA Reijneveld)

Mafirakureva N
Health economics of blood transfusion in Zimbabwe
(prof MJ Postma, dr M van Hulst, dr S Khoza)

Mapako T
Risk modelling of transfusion transmissible infections
(prof MJ Postma, dr M van Hulst)

Heuvel M van den
Developmental and behavioral problems in pediatric care; early identification
(prof SA Reijneveld, dr DEMC Jansen, dr BCT Flapper)

Bonvanie IJ
Functional somatic symptoms in adolescence and young adults; personal vulnerabilities and external stressors
(prof JGM Rosmalen, prof AJ Oldehinkel, dr KAM Janssens)

Greeff JW de
Physically active academic lessons: effects on physical fitness and executive functions in primary school children
(prof C Visscher, prof RL Bosker, dr E Hartman, dr S Doolaard)

For more 2016 and earlier theses visit our website.

Acknowledgements

First of all, I would like to praise the almighty Allah for giving me the strength and opportunity to bring this PhD work to an end. There are lots of things that I have learnt during the trajectory of my PhD research, in which I have to thank many intellectual people. It is a great opportunity to convey my deepest indebtedness to my supervisorial team Prof. dr. Edwin R. van den Heuvel, Eindhoven University of Technology, Eindhoven, the Netherlands, dr. Richard Bruggeman and dr. Behrooz Z. Alizadeh, University Medical Center Groningen (UMCG) for their constant encouragement and invaluable personal guidance for carrying out my thesis. I express my deep feelings and gratitude to them not only for supervising my work but also for many more helpful suggestions they extended me without which it would have not been possible for me to complete my research.

Dear Edwin, my esteemed promoter, I would like to express my sincere gratitude for accepting me as a PhD student under your supervision. I still remembered the day of interview where I get interviewed with another project but two weeks later you offered me to work on GROUP project. Thank you for giving me the chance. You gave me ample opportunities to discuss my works, new statistical methodology, new ideas when you were in UMCG and later in TU/e. You have showed your tremendous support and guidance despite your busy schedule. I found you not only a good supervisor and researcher but also an honest man. Thank you for being my promoter.

Dear Richard, I would like to extend my best thanks for selecting me as a PhD candidate to work within the GROUP-project. I remembered the day when I came to UMCG for the first time to give a presentation for my candidacy in front of you and Edwin. I was really happy when both of you offered me for the position. You gave lots of ideas related to my research in due course, which helped me a lot to think about new research lines especially in neurocognition. Moreover, I really enjoyed in our one-to-one regular meeting about statistics and psychiatry related issues. You are one the kindest person that I ever met, you gave me a lot of opportunities to access your office without any appointment and discuss my issues with you. I greatly appreciate your excellent assistance regarding all official letters whenever I asked for. You are very much helpful and supportive for me in both academic and non-academic (personal) work. Moreover, I would also like to thank you for translating my thesis summary into Dutch.

Dear Behrooz, I would like to express my deep gratitude to you for untiring patience and attention given to me whilst completing this thesis. I am really happy to find you as my daily supervisor. I learnt a lot from you about many topics of epidemiology and genetic epidemiology and still learning. You showed me many genetic data analysis software programs, which I am using now a days a lot. I am grateful to you for encouraging and supporting me in pursuing my ideas. I could not remember how many times I access your office without appointment and you always welcomed me with smiley words “hey Atique, how are you”. You always showed me the direction of my future work. You are also very supportive for my work at UMCG and I feel very happy in our MDL unit and personal one-to-one meeting. Thank you for everything.

I would like to show my appreciation to the assessment committee Prof. dr. E.C. (Ernst) Wit, Prof. dr. A.J. (Tineke) Oldehinkel and Prof. dr. A.C. (Lydia) Krabbendam for providing valuable comments and suggestions.

I would like to extend my cordial thanks to all the members in my research team and co-authors who worked with me along the way. It was my privilege to work with Feroz Hayat Khan, Sanne Smith-Apeldoorn, dr. Annemarie Stiekema, dr. Piotr Quee, Frank, Tesfa, Stefanie, dr. Lisette van der Meer, dr. Agna A. Bartels-Velthuis, dr. Edith Liemburg, dr. Jojanneke Bruins, dr. Anne Crijns, Hans Burgerhof, dr. Talitha Feenstra, dr. Ilja Nolte, Prof. dr. Bob Wilffert, Prof. dr. Hans Langendijk, Prof. dr. Marieke Pijnenborg, Prof. dr. André Aleman, Prof. dr. Harold Snieder, Prof. dr. Lieuwe de Haan and Prof. dr. Jim van Os.

Dear Prof. dr. Eelko Hak, I owe a deep sense of gratitude for giving me the chance of finishing my thesis during the work of UNISEC project. I really appreciate your concern for my well-being. I would also like to thank Nynke, Thea, Eva, Ivan, Ira, Sylvi, Jannie and others for your valuable suggestions during my work in PharmacoEpidemiology department.

I am very grateful to the RGOc, Department of Psychiatry and Epidemiology where I received excellent academic and social support. For the awesome people in the department, thank you for the company and memories: Sjoerd, Lian, Piotr, Ellen Visser, Rob, Frederike, Kaying, Jan, Fionneke, Ricardo, Maritt, Jaap Janssen and all others. I would like to thank our efficient ladies Martha Messchendorp, Margo Jongsma, and Esther Hollander-Westera at the Department of Psychiatry, Aukje van der Zee and Roelian Geuze at the Department of Epidemiology for administrative support. Dear Martha and Margo, from beginning to till now, you helped me a lot. Thank you so much. I would also like to thank all the members of RGOc to whom I worked with. Dear Kaying and Jan, thank you so much for correcting grammatical issues in my discussion part. Dear Sanne, thank you for polishing the Dutch summary. I am also very thankful to Thembile, Nazanin and Nino. I express my thanks to all members of MDL unit of Epidemiology specially Marzyeh, Mahmud, Shifteh, Tesfa, Reza, Fareeba, Elnaz, Vera, Kim and others for sharing scientific knowledge and arranging social events.

Dear Lian and Piotr, I desire to express my special thanks both of you. Before coming to Groningen from Bangladesh, I had a housing problem. You guys continued to contact me even at the time of take-off the plane from Bangladesh, and informed me that everything will be fine. Indeed, everything went well. Nadine and you guys managed everything nicely. Thanks also to Nadine. Beside theses, we together shared the same room and discussed many things which I liked most.

I am grateful for the generous time and effort offered by the patients, their families and healthy subjects who make the GROUP project possible. I would like to thank all research personnel involved in the GROUP project especially Joyce van Baaren and Erwin Veermans. Dear Erwin, whenever I faced any difficulty in data management issues, you always helped me and explained them to me clearly. Thanks for everything.

My paranymphs, Sanne and Jan, thank you for the great help in organizing my thesis defense. I am really grateful that you are willing to help me with everything despite your tight schedule.

Dear Rana bhai and Mithila vabi, I cannot thank you enough for the support and help that you gave me during my stay in Groningen. You helped me tremendously to make me feel comfortable here in Groningen. I had a great time with you. In fact you made my life in Groningen easier. I disturbed you a lot and sorry for that.

I express my thanks to Dr. Feroz Hayat Khan bhai for your constant help. Indeed, we had a great time here when you were in Groningen. Many thanks go to Omar Faruque bhai for supporting me and I wish the very best of luck for your PhD. I would like to thank all of my other Bangladeshi and Indian friends, especially Khokon bhai, Nasrin vabi, Asad bhai, Muztafa bhai, Islam bhai, Maruf, Tushar-Jasmine, Soumen-Soutri, Shiva and many more whom I met in Groningen.

Romel Bhuia (RB! and president of WGB! You know what I meant ☺), thanks my friend, you always encouraged and supported me a lot when I discussed or shared anything with you. During my PhD time, I called you countless time and you listened to my stubborn things and always told me that "I will surely manage it". Now is the time for your success. I wish you all the best for your PhD. I also have to thank my all well-wishers, especially my friends Dr. Arif, Mamun, Masud, Tutul, Faruk, Amit, Rama, Mamun (Stat), Mahbub, Gunda, Salman, Sujoy, Shahed, Kawsar, Biplob, Humaun Kabir Ansary (thanks for the amazing tour of Niagara Falls!), Leader Kabir, Dr. Golam Sarwar (Sohan), Wahid, Rajib, Feroz Khan (Leton), Khaled bhai, Patwary bhai and many more for their co-operation and support.

I am very grateful to Rajwanur Rahman for enormous support during my PhD time. Dear Mith, I bothered you a lot and I knew that but you never told me "NO". Whenever I encountered any problem with programming, I asked you "Mith, I want this right now" and you provided me the solutions instantly. I believe without your help, the completion of this thesis would have been difficult. Of course, you deserve another thanks for managing my all accounts ☺.

Dear Shopon bhai (Dr. Lutful Alahi Kawsar), I am really lucky to get you as my elder brother and colleague. I always asked for your help either on official purpose or personal issues and you supported me a lot. Thank you for your wonderful support. I also desire to express my big thanks to Dr. Khalid bhai, Dr. Ohid bhai, Zia bhai, Dr. Monsur bhai, Dr. Jamal, Dr. Sumonkanti Das and all my charming junior colleagues at the department of Statistics, SUST for good suggestions.

Dear Prof. Md. Ashraf-Ul-Alam sir, I would like to convey my deepest indebtedness to you for your invaluable personal guidance from very beginning of my study in Statistics to till date. During my tough and difficult times, you inspired me a lot and provided me a good direction. Thank you so much for your supportive help.

I express my profound acknowledgement to my employer in Bangladesh (Department of Statistics, Shahjalal University of Science and Technology, Sylhet-3114) for providing me study leave to pursue the PhD research. I would like to thank the current and ex-dean and head of the department of Statistics, Prof. Dr. Ahmad Kabir, Prof. Dr. Kabir Hossain, Prof. Dr. Zakir Hossain and Prof. Dr. Sabina Islam for their support and priceless help on several official letters of study leave. I am also grateful to all of my other respected teachers and staffs of the Department of Statistics, SUST, Sylhet, Bangladesh.

Now time to thank my wonderful family. This thesis would not have been possible without the love of family. Therefore, I would like to express my heartiest honour to my parents, parents in law and uncle Md. Ataur Rahman whose affections, moral supports and encouragements enabled me to finish this research work. All the credit goes to my parents for what I am today. Also, I find no way to express my feeling of sincere gratitude to my grandfather Late Tozammel Haque (Daroga), from whom I got affections and inspiration since my early education. I am also thankful to my other family members, especially my adorable brothers (Robi and Rifat), sisters (Bonani, Babita and Rani), brothers in law (Kazi Rafiqul Islam and Rezaul Karim Khan), siblings in law (Tasnia and Shadman), nephew (Shammo), nieces (Munni and Moumi), uncles, aunts and cousins for their continuous inspiration during the period of my study. Dear Robi and Rifat, you deserved special thanks because of taking care of our family in absence of mine.

Most of all, I am deeply indebted to my beloved wife Dr. Rukhsana Ferdous Rushmi for her continuous encouragement, unconditional support and sacrifices throughout the course of research. Your patience and understandings during the period of my study allowed me to spend most of the time on this thesis. Thank you for supporting me for everything. Last but not least, I have to acknowledge the most important person in my life, my lovely sweet daughter Yariqa Sharin for abiding my ignorance and the patience she showed during the last two years of my PhD trajectory. Besides, Yariqa, you bring so much joy and happiness in our life. There is nothing more satisfying when I see your happy and smiling face.

সবাইকে অনেক ধন্যবাদ।

ABOUT THE AUTHOR

Md. Atiqul Islam was born on January 01, 1981 in Pabna, Bangladesh, where he obtained secondary and higher secondary school certificates accredited by Santhia Pilot High School and Govt. Edward College. Next, he studied both Bachelor of Science (B.Sc.) and Master of Science (M.Sc.) programs in Statistics at the Department of Statistics, Shahjalal University of Science and Technology (SUST), Sylhet, Bangladesh from August 1999 to November 2006. He secured first position in both programs and awarded in recognition of the highest talent and greatest achievement among all graduates of the Department of Statistics at the Bachelor's and the Master's level examinations. After graduating as a statistician, he was appointed as a young academic lecturer at the same department. Later in 2009, he received VLIR-UOS awards scholarships from Flemish Government to pursue Master in Biostatistics at the Center for Statistics (CenStat), Hasselt University, Belgium. This has led to his master thesis in Biostatistics, entitled "*Herd-level risk factors associated with bovine brucellosis sero-positivity and abortion in Banglades*h". He applied joint mixed effects modeling to identify the factors that associated with bovine brucellosis sero-positivity and abortion. In 2011, he obtained his second Master in Biostatistics from CenStat.

It was November 2011, Atiqul was interviewed and granted for a PhD position jointly supported by the Department of Psychiatry and Department of Epidemiology, University Medical Center Groningen (UMCG), University of Groningen, the Netherlands. During his PhD trajectory, he explored the heterogeneity in cognitive functioning and clinical symptoms in people with psychotic disorders and their unaffected siblings using cross-sectional and longitudinal data under Genetic Risk and Outcome of Psychosis (GROUP) project. Throughout this period, he was also involved in teaching and supervising Medical and Clinical and Psychosocial Epidemiology students. He is also appointed as a researcher in data handling, statistical and genetic analysis of GROUP-project till date.

Since September 2016, Atiqul is working as a post-doctoral researcher in Biostatistics at the Department of PharmacoTheraphy, -Epidemiology & Economics, University of Groningen. He also holds the position as an Assistant Professor at the Department of Statistics, Shahjalal University of Science and Technology, Sylhet-3114, Bangladesh.